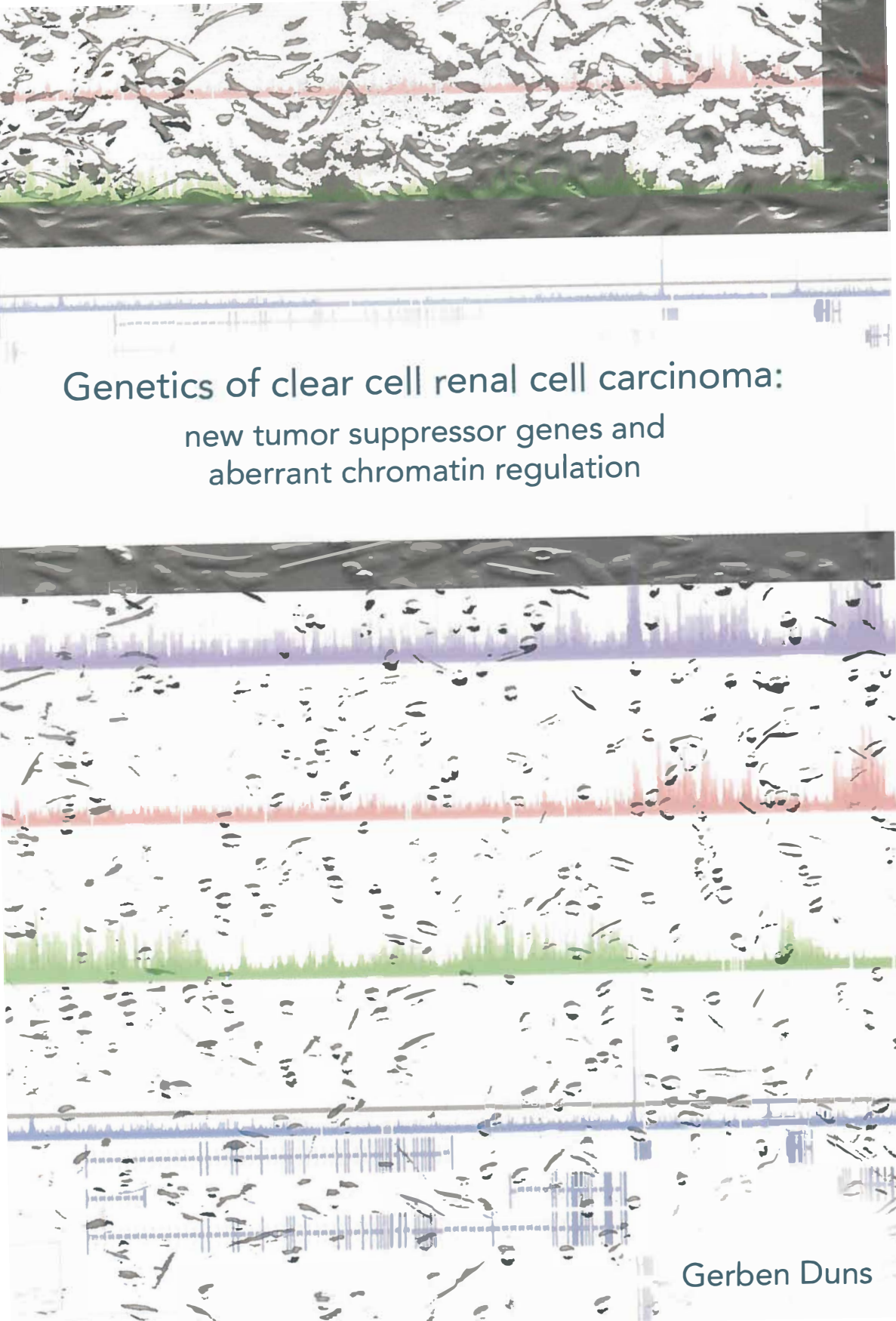


Genetics of clear cell renal cell carcinoma:

new tumor suppressor genes and aberrant chromatin regulation

Gerben Duns

Genetics of clear cell renal cell carcinoma: new tumor suppressor genes and aberrant chromatin regulation

Gerben Duns

Stellingen behorende bij het proefschrift

Genetics of clear cell renal cell carcinoma:
new tumor suppressor genes and aberrant chromatin regulation

1. De 3p21 regio bevat met *PBRM1*, *SETD2* en *BAP1* meerdere ccRCC- geassocieerde tumorsuppressor- genen (*dit proefschrift*).

2. Verschillende ccRCC tumoren hebben inactiverende mutaties in zowel *VHL* als *PBRM1* als *SETD2*, wat erop wijst dat inactivatie van deze genen individueel niet voldoende is voor de ontwikkeling van ccRCC tumoren (*dit proefschrift*).

3. *PBRM1* is in een meerderheid van ccRCC tumoren geïnactiveerd en is daarmee een interessant aangrijpingspunt voor nieuw te ontwikkelen gerichte therapieën tegen ccRCC (*dit proefschrift*).

4. Een verstoring van de regulering van chromatine is een essentiële gebeurtenis in de ontwikkeling van ccRCC tumoren (*dit proefschrift*)

5. Inactivatie van *VHL* alleen leidt niet tot de ontwikkeling van ccRCC tumoren. Het is dan ook een logische stap om de (ccRCC) tumoren van patiënten die leiden aan VHL-syndroom te screenen op mutaties in de "nieuwe" ccRCC- geassocieerde tumorsuppressor-genen.

6. Het doen van (een promotie-)onderzoek is te vergelijken met het maken van een wielertocht in de bergen: je vraagt je geregeld af waarom je er aan begonnen bent, maar het behalen van resultaat/ bereiken van de top geeft voldoening.

7. " De weereld is een speeltooneel, elck speelt zyn rol en kryght zijn deel" (*Joost van den Vondel*).

8. Voetballers die bij clubs spelen die overheidssubsidies ontvangen zouden niet meer moeten verdienen dan de Balkenende- norm.

9. De doodstraf is een wrede en onmenselijke straf, net zoals de in Nederland bestaande levenslange gevangenisstraf zonder enig uitzicht op vervroegde vrijlating.

Gerben Duns, 7 december 2011

Gerben Duns

Genetics of clear cell renal cell carcinoma: new tumor suppressor genes and aberrant chromatin regulation

Thesis University of Groningen with summary in Dutch

The studies designed in this thesis were financially supported by the Dutch Cancer Society (Grant 2007-3892), the Foundation "De Drie Lichten" in the Netherlands, the "Jan Kornelis The Cock" Foundation, and the René Vogels Stichting

Printing of this thesis was financially supported by the Dutch Cancer Society, Rijksuniversiteit Groningen, and Groningen University for Drug Exploration (GUIDE)

Cover design by Ivo Tanis (ivotanis@gmail.com)

Lay-out by Off Page, Amsterdam

Printed by Off Page, Amsterdam

ISBN: 978-90-367-5193-3

Cmb.

RIJKSUNIVERSITEIT GRONINGEN

Genetics of clear cell renal cell carcinoma: new tumor suppressor genes and aberrant chromatin regulation

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
woensdag 7 december 2011
om 14:30 uur

door

Gerben Duns
geboren op 14 oktober 1981
te Apeldoorn

Promotor:	Prof. dr. R. M. W. Hofstra
Copromotores:	Dr. K. Kok Dr. E. van den Berg- de Ruiter
Beoordelingscommissie:	Prof. dr. A. H. M. Geurts van Kessel Prof. dr. S. J. M. Jones Prof. dr. J. M. Nijman

Paranimfen: Annemieke T. van der Goot
Willem M. Kok

CONTENTS

GENERAL INTRODUCTION

1

GENERAL INTRODUCTION

Clear cell renal cell carcinoma

Renal cell carcinoma (RCC) is the most common type of kidney cancer in adults, and accounts for 3% of all new cancers diagnosed in the Western world (Jemal et al., 2010). The incidence of RCC is the lowest in Asia and Africa (with the lowest recorded rates less than 1 new annual case per 100,000 population, age standardized) and the highest in Europe and North America (with the highest recorded rates in both men and women observed in the Czech Republic, with 20.2 and 9.5 annual new cases per 100,000, respectively). The incidence in the Netherlands is 7.9 and 4.5 per 100,000 in men and women, respectively (Curado et al., 2007).

There is evidence, supported by animal experiments and histological characterization of pre- and early stages of epithelial renal tumors that RCC derives from the mature nephron (Motzer et al., 1996). RCC is a clinicopathologically heterogeneous disease that consists of a number of histologically defined entities. The present classification of renal cell carcinoma was established by international consensus and led to the World Health Organization (WHO) 2004 classification. According to this classification, which is based on both morphological and genetic characteristics, RCC can be grouped into clear cell, papillary, chromophobe, collecting duct, medullary carcinoma, and unclassified categories (Eble et al., 2004). The WHO classification of RCC and the cytogenetic aberrations associated with the different RCC subtypes are summarized in Table 1. Clear cell renal cell carcinoma (ccRCC) is by far the most prevalent subtype,

Table 1. Classification renal cell carcinoma (RCC) subtypes.

Histologic subtype (WHO classification 2004)	Prevalence (%)	Cytogenetic findings
Clear cell RCC	70	3p deletions, *VHL* mutations
Papillary RCC	10 to 15	Trisomy of chromosomes 7 and 17, loss of Y chromosome
Chromophobe	4 to 6	Loss of multiple chromosomes: 1,2,6,10,13,17,21
Hereditary cancer syndromes	5	Dependent on subtype
Collecting duct carcinoma	<1	Loss of multiple chromosomes: 1,6,14,15,22; gain of chromosome 3
Medullary carcinoma	<1	Extracellular matrix gene loss
Mucinous tubular and spindle cell carcinoma	<1	Loss of multiple chromosomes: 1,4,6,8,13,14
Neuroblastoma-associated RCC	<1	Multiple gene loss (most often 20q13)
Xp11.2 translocation-TFE3 carcinoma	<1	Translocations involving Xp11.2
Unclassified lesions	4	

Sources: Eble et al., 2004; Prasad et al., 2006

comprising more than 70% of adult malignant kidney tumors. ccRCC is a malignant tumor composed of cells with a clear cytoplasm due to an intensive intracytoplasmatic accumulation of glycogen and lipids. Electron-microscopic and immunohistochemical features resembling the proximal tubule are seen in ccRCC, suggesting that ccRCC tumors are derived from the epithelial cells of the proximal tubulus (Thoenes et al., 1985). Figure 1 shows a schematic picture of the different parts of the nephron from which ccRCC and the other RCC subtypes are thought to be derived from (Prasad et al., 2006).

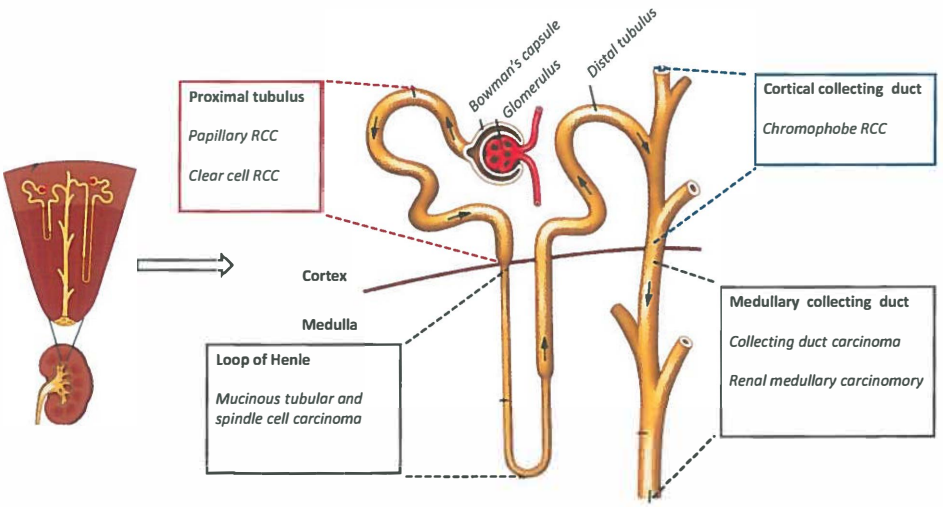


Figure 1. Schematic overview of the nephron, showing the different parts from which the RCC subtypes are thought to be derived from. *Figure is modified from http://www.beltina.org.*

In general ccRCC tumors occur sporadically. The average age of diagnosis is 60-64 years. However, 7% of sporadic cases of ccRCC are diagnosed in patients who are younger than 40, and rare cases have been described in patients between 14 and 18 years, without evidence of familial disorders (Bruder et al., 2004). The three most well-established risk factors associated with the development of ccRCC are smoking, obesity and hypertension (Chow et al., 2000; McGuire et al., 2011; Parkin et al., 1985).

Hereditary clear cell renal cell carcinoma

Approximately 2% of ccRCC tumors are familial (reviewed by Maher, 1996). The most frequent cause of familial RCC is a dominantly inherited cancer syndrome named VHL disease, which is estimated to occur at rates of 1:45500 to 1:100000 (Maddock et al., 1996; Maher et al., 1990a). VHL disease is the result of germline mutations in the *Von Hippel Lindau* (*VHL*) gene, and is characterized by the development of benign and malignant tumors across several organs, mainly in parts of the body with a high density

of blood vessels (reviewed by Kaelin, 2002). The lifetime risk of developing one or more ccRCC tumors in VHL disease is over 70% by the age of 60 (Maher et al., 1990b). In a majority of sporadic ccRCC tumors VHL is inactivated as well (Gnarra et al., 1994; Nickerson et al., 2008).

An intriguing finding has been the frequent occurrence of translocations involving chromosome 3 in familial cases of ccRCC that are not linked to VHL. The best known case is undoubtedly the t(3,8)(p14;q24) translocation reported by Cohen et al. (1979) in a three-generation family in which all members with bilateral ccRCC carried this constitutive translocation. The chromosome 3 breakpoint appeared to map in intron 3 of *FHIT* (Zimonjic et al., 1997). Subsequently, several other familial cases of ccRCC have been identified with a constitutive translocation involving chromosome 3 (Bonne et al., 2007; Foster et al., 2007; Rodríguez-Perales et al., 2004). The various cases, however, do not appear to share a breakpoint on chromosome 3. And, although detailed analysis of a number of these translocations has resulted in the identification of genes disrupted by the translocation, none of these genes appear to be mutated in a substantial number of sporadic cases of ccRCC. It has been suggested that in these cases susceptibility to ccRCC may result from chromosome 3 instability, predisposing to loss of 3p. Indeed, analysis of the tumors in these families often revealed loss of the translocation chromosome that contained the short arm of chromosome 3 (reviewed by Bodmer et al., 2002; Li et al., 1993).

Therapy

In most cases, ccRCC tumors are still localized at the moment they are found. Surgical removal is the standard of care for these tumors, and offers long-term recurrence-free survival, although for patients who suffer from metastatic disease or develop distant relapse, the scenario is worse. Renal cell carcinomas are resistant to traditional cytotoxic chemotherapy. Immunotherapy, which has been the most frequently used therapy until recently, has not been satisfactory in most cases either (reviewed by Bukowski et al., 2011).

Advances in the understanding of ccRCC biology have resulted in the development of targeted approaches, which include the use of humanized monoclonal antibodies against VEGF, small-molecule tyrosine-kinase inhibitors, and inhibitors of mTOR. These approaches, which target pathways that are characteristic and essential for ccRCC tumors, show superior response rates compared to conventional strategies (reviewed by Clark, 2009).

Genetics of clear cell renal cell carcinoma

At the genomic level, ccRCC is characterized by a range of copy number alterations, of which loss of the short arm of chromosome 3 is by far the most frequent (Van den Berg et al., 1993). Genomic regions with a high incidence of allelic loss are thought to harbor so-called tumor suppressor genes (TSGs), defined by Haber and Harlow (1997) as "genes that sustain loss-of-function mutations in the development of cancer". The allelic losses are regarded as one of the two genetic alterations that are necessary to biallelically inactivate TSGs (Knudson, 1971). In ccRCC a tumor suppressor gene

has been identified in the recurrently lost 3p25 region, namely *VHL*. One allele of the VHL gene is thus deleted in most - if not all - ccRCCs. The remaining allele is inactivated by small mutations in 43–82% of ccRCC samples (Foster et al., 1994a; Gnarra et al., 1994; Nickerson et al., 2008; Shuin et al., 1994). In combination with the predisposition to multicentric ccRCC in patients with VHL germline mutations, these data have pinpointed VHL as a classic tumor suppressor gene. Further support came from the observation that reintroduction of wildtype *VHL* into *VHL*-null ccRCC cells suppresses tumorigenicity both *in vitro* and *in vivo* (Chen et al., 1995; Gnarra et al., 1996; Iliopoulos et al., 1995; Schoenfeld et al., 1998).

As mentioned, in a subset of ccRCC tumors, *VHL* is not biallelically inactivated. Thus, at least for these tumors, *VHL* inactivation is probably not the driving force for the observed 3p loss. Moreover, *VHL* inactivation alone causes a senescence-like phenotype *in vitro* and *in vivo* (Young et al., 2008), and although kidneys from patients suffering from *VHL* disease contain *VHL* mutations in hundreds of pre-neoplastic lesions, only a few of these are thought to develop into ccRCC tumors (Mandriota et al., 2002). These observations strongly suggest that alternative and/or additional ccRCC cancer genes play a role in ccRCC development. The observed frequent loss of regions of 3p, other than 3p25, in ccRCC tumors strongly suggests that these regions might contain unidentified ccRCC tumor suppressor genes. Many studies have therefore attempted to precisely map the position and the boundaries of the deleted regions of 3p.

Loss-of-heterozygosity (LOH) studies suggest that 3p21 loss is slightly more frequent than, and/or may occur separately from, loss of the 3p25 and 3p14 loci (Clifford et al., 1998; Van den Berg et al., 1997). Allelic loss of 3p25 or of 3p12-14, but not of 3p21, has been shown to occur in renal cell adenoma, suggesting that loss of the 3p21 region is a prerequisite for malignant development of ccRCC (Van den Berg et al., 1998). This suggests that loss of these 3p segments may independently contribute to ccRCC development, and indicates the presence of tumor suppressor genes at the 3p21 and 3p14 regions. A few genes located at these regions on 3p have been suggested to have a tumor suppressor role in ccRCC, most notably *fragile histidine triad protein gene (FHIT)*, and *ras association (RalGDS/AF-6) domain family member 1 (RASSF1A)* (*Figure 2*). *FHIT* co-localizes with the fragile site FRA3B at 3p14.2. A multitude of studies in different types of tumors have shown various molecular alterations, including CpG-island hypermethylation, leading to inactivation of FHIT. FHIT participates in a number of important signal pathways in response to endogenously or exogenously applied stressful agents, particularly those that cause DNA damage (reviewed by Saldivar et al., 2010). There are still questions to pursue concerning the selective advantage conferred to cells by loss of expression of FHIT, and the involvement of FHIT in ccRCC is disputed (Bugert et al., 1997 Van den Berg et al., 1997).

RASSF1, located at 3p21, was initially suggested to be a tumor suppressor gene in lung cancer, based on its location in a region of frequent LOH and the absence of expression as a result of promoter hypermethylation of the remaining allele (Agathanggelou et al., 2001). Dreijerink et al. (2001) showed promoter

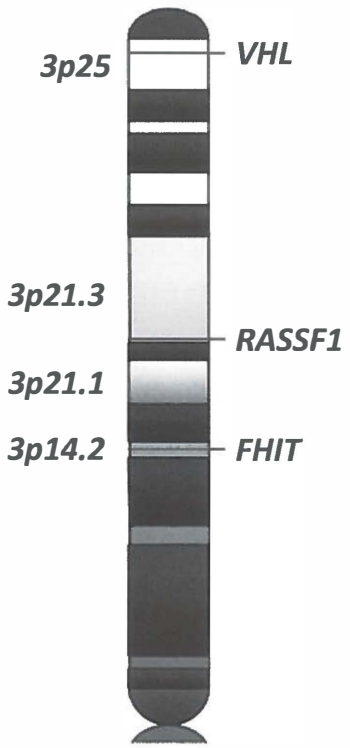


Figure 2. Schematic picture of the short arm of chromosome 3, showing the locations of the *VHL* ccRCC tumor suppressor gene, and the putative ccRCC tumor suppressor genes *RASSF1* and *FHIT*.

hypermethylation of *RASSF1* in 39 out of 43 primary ccRCC tumors. The *RASSF1* gene has two GC-rich promoters that produce two isoforms, RASSF1A and RASSF1C. In ccRCC, both these transcripts suppress the tumorigenicity of the KRC/Y RCC cell line in SCID mice (Li et al., 2004). This can be regarded as a hallmark of a tumor suppressor gene. However, as *RASSF1a* is not inactivated by mutation but by promoter methylation, this gene may not be seen as a tumor suppressor gene in the classic sense. The aberrant promoter hypermethylation is a secondary effect as another event has to precede this one in order to establish the aberrant hypermethylation. Moreover, the process of hypermethylation will simultaneously affect all alleles of a gene in a single event. This is in contrast with the situation for the classic tumor suppressor genes where inactivation requires two independent events (Knudson et al., 1971). Therefore, inactivation of *FHIT* and *RASSF1* is probably not the driving force for the frequently observed loss of 3p14 and 3p21, respectively.

The proposed model, suggesting that loss of 3p segments occurs independently in ccRCC tumorigenesis, is challenged by recent studies that show that for the majority of ccRCCs the loss of 3p most likely encompasses the entire p-arm. In a study by Sükösd et al. (2003), based on microsatellite analyses, 95 out of 98 ccRCC tumors showed loss of a segment ranging from pter to 3p12. In a SNP-based study by Toma et al. (2008), 15 out of 20 tumors had lost the entire p-arm. Neither of these studies reported any cases with retention of heterozygosity at the VHL locus, while retention at the *FHIT* locus was observed in six cases in the combined studies. The data of Braga et al. (2002) and Chudek et al. (1997) underscore these conclusions. Their data on 55 and 100 ccRCCs, respectively, indicate large terminal deletions for 42 and 94 cases, respectively, of which 28 and 89 most likely encompass almost the entire p-arm. Chudek et al. (1997) reported retention of the most telomeric part of 3p in 6 cases, but this segment did not include the VHL locus. Many studies that report interstitial 3p deletions in ccRCC tumors are based on the analysis of microsatellite markers. The interpretation of microsatellite data of primary tumors is complicated to say the least. Primary tumors have an admixture of normal tissue, and this will dampen

the effect of LOH on the observed ratios between the alleles. A study by Liu et al. (1999) suggested that up to 50% of the low molecular weight allele deletions in tumor samples may go undetected in this type of study. Thus, reported interstitial deletions, or interruptions of regions of deletion, based on only one marker should be regarded with caution. Alimov et al. (2000) used an analysis protocol that took the findings of Liu et al into account, and did report interstitial deletions in six primary RCCs. Wilhelm et al. (1995) avoided the influence of contaminating normal cells by using short-term cultures of primary ccRCC. All 41 tumors that they analyzed appeared to have strictly terminal deletions of the major part of 3p, including the FHIT locus. Thus, it cannot be excluded that interstitial deletions do occur, but they are presumably very scarce.

The combined genomic studies strongly suggest that the short arm of chromosome 3 may harbor several, as yet unidentified, tumor suppressor genes that contribute to ccRCC development. Identifying ccRCC tumor suppressor genes by positional cloning strategies alone has proven extremely difficult. Decades of research on the identification of the tumor suppressor genes other than *VHL* that underlie the high frequency of 3p deletions in ccRCC has not led to the identification of a clear-cut tumor suppressor gene. The identification of tumor suppressor genes was hampered by the omission of methods to screen large numbers of genes in parallel for small (point) mutations. This situation has changed in the last decade. New methods like Gene Identification by Nonsense-mediated mRNA decay Inhibition (GINI), and, more recently, next-generation sequencing-based approaches, have given us new and powerful tools to identify tumor suppressor genes.

AIM

The aim of this thesis was to identify new tumor suppressor genes that are involved in the development of clear cell renal cell carcinoma (ccRCC). Subsequent detailed functional analysis of these genes should help us to understand the initial processes that underlie the development of ccRCC.

OUTLINE

In ***Chapter 2***, we applied *Gene Identification by Nonsense mediated mRNA decay Inhibition (GINI)* in the search for novel ccRCC tumor suppressor genes. This led to the identification of *SETD2* as a new candidate ccRCC tumor suppressor gene. Using the current knowledge on ccRCC genetics, we evaluated the use of *GINI* in the search for tumor suppressor genes in ***Chapter 3***.

SETD2 codes for a histone methyltransferase that is nonredundantly responsible for trimethylation of lysine 36 of histone H3 (H3K36me3). This modification has been suggested to have a function in alternative splice site choice. In ***Chapter 4***, we describe the initial steps that should result in the elucidation of the mechanisms through which *SETD2* inactivation could contribute to ccRCC development. We investigated the

effect of siRNA-mediated *SETD2* knockdown in proximal tubular epithelial cells on transcript levels and alternative splice site choice.

The recently developed next-generation sequencing (NGS) technologies make it possible to sequence large numbers of genes in parallel. In ***Chapter 5***, we continued our search for ccRCC tumor suppressor genes by sequencing all coding regions and exon-intron boundaries from the short arm of chromosome 3 in ten primary ccRCC tumors, using exome capturing followed by NGS. This led to the identification of *PBRM1* as a candidate ccRCC tumor suppressor gene.

Besides protein coding genes, non-coding transcripts, including microRNAs, have been shown to have a function in the development of several tumor types. Therefore, we compared miRNA transcript levels in ccRCC-derived cell lines with the miRNA levels in normal proximal epithelium, and correlated these transcript levels with copy numbers and transcript levels of predicted targets, as described in ***Chapter 6***.

In ***Chapter 7***, the work described in this thesis is discussed.

REFERENCES

Agathanggelou A, Honorio S, Macartney DP, et al. Methylation associated inactivation of RASSF1A from region 3p21.3 in lung, breast and ovarian tumours. Oncogene 2001;20:1509.

Alimov A, Kost-Alimova M, Liu J, et al. Combined LOH/CGH analysis proves the existence of interstitial 3p deletions in renal cell carcinomas. Oncogene 2000;19:1392.

Bonne A, Vreede L, Kuiper RP, et al. Mapping of constitutional translocation breakpoints in renal cell cancer patients: identification of KCNIP4 as a candidate gene. Cancer Genet Cytogenet 2007;179:11.

Bodmer D, Van den Hurk W, Van Groningen JJ, et al. Understanding familial and non-familial renal cell cancer. Hum Mol Genet 2002;11:2489.

Braga E., Senchenko V, Bazov I, et al. Critical tumor-suppressor gene regions on chromosome 3P in major human epithelial malignancies: allelotyping and quantitative real-time PCR. Int J Cancer 2002;100:534.

Bruder E, Passera O, Harms D, et al. Morphologic and molecular characterization of renal cell carcinoma in children and young adults. Am J Surg Pathol 2004;28:1117.

Bugert P, Wilhelm M, Kovacs G. FHIT gene and the FRA3B region are not involved in the genetics of renal cell carcinomas. Genes Chromosomes Cancer 1997;20:9.

Bukowski RM. Cytokine therapy for metastatic renal cell carcinoma. Semin Urol Oncol 2011;19:148.

Chen F, Kishida T, Duh FM, et al. Suppression of growth of renal carcinoma cells by the von Hippel-Lindau tumor suppressor gene. Cancer Res 1995;55:4804.

Chow WH, Gridley G, Fraumeni JFJr, et al. Obesity, hypertension, and the risk of kidney cancer in men. N Engl J Med 2000;343:1305.

Chudek J, Wilhelm M, Bugert P, et al. Detailed microsatellite analysis of chromosome 3p region in non-papillary renal cell carcinomas. Int J Cancer 1997;73:225.

Clark PE. The role of VHL in clear-cell renal cell carcinoma and its relation to targeted therapy. Kidney Int 2009:76;939.

Clifford SC, Prowse AH, Affara NA, et al. Inactivation of the von Hippel-Lindau (VHL) tumour suppressor gene and allelic losses at chromosome arm 3p in primary renal cell carcinoma: evidence for a VHL-independent pathway in clear cell renal tumorigenesis. Genes Chromosomes Cancer 1998;22:200.

Cohen AJ, Li FP, Berg S, Marchetto DJ, et al. Hereditary renal-cell carcinoma associated with a chromosomal translocation. N Engl J Med 1979;301:592.

Curado. M. P., Edwards, B., Shin. H.R., et al. Cancer Incidence in Five Continents, Vol. IX. IARC Scientific Publications 2007;160.

Dreijerink K, Braga E, Kuzmin I, et al. The candidate tumor suppressor gene, RASSF1A, from human chromosome 3p21.3 is involved in kidney tumorigenesis. Proc Natl Acad Sci U S A 2001;98:7504.

Eble JN, Sauter G, Epstein JI, et al. World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs. IARC Press: Lyon 2004.

Foster RE, Abdulrahman M, Morris MR, et al. Characterization of a 3;6 translocation associated with renal cell carcinoma. Genes Chromosomes Cancer 2007;46:311.

Foster K, Prowse A, Van den Berg A, et al. Somatic mutations of the von Hippel-Lindau disease tumour suppressor gene in non-familial clear cell renal carcinoma. Hum Mol Genet 1994;3:2169.

Gnarra JR, Tory K, Weng Y, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. Nature Genet 1994;7:85.

Gnarra JR, Zhou S, Merril MJ, et al. Post-transcriptional regulation of vascular endothelial growth factor mRNA by the product of the VHL tumor suppressor gene. Proc Natl Acad Sc USA 1996;93:10589.

Haber D, Harlow E. Tumour-suppressor genes: evolving definitions in the genomic age. Nat Genet 1997;16:320.

Iliopoulos O, Kibel A, Gray S, et al. Tumour suppression by the human von Hippel-Lindau gene product. Nat Med 1995;1:822.

Jemal A, Siegel R, Xu J, et al. Cancer Statistics, 2010. CA Cancer J Clin 2010;60:277.

Kaelin Jr. WG. Molecular basis of the VHL hereditary cancer syndrome. Nat Rev Cancer 2002;2:673.

Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68:820.

Li FP, Decker JH, Zbar B, et al. Clinical and genetic studies of renal cell carcinomas in a family with a constitutional chromosome 3;8 translocation. Genetics of familial renal carcinoma. Ann Intern Med 1993;118:106.

Li J, Wang F, Protopopov A, et al. Inactivation of RASSF1C during in vivo tumor growth identifies it as a tumor suppressor gene. Oncogene 2004;23:5941.

Liu J, Zabarovska VI, Braga E, et al. Loss of heterozygosity in tumor cells requires re-evaluation: the data are biased by the size-dependent differential sensitivity of allele detection. FEBS Lett 1999;462:121.

Maddock IR, Moran A, Maher ER, et al. A genetic register for von Hippel-Lindau disease. J Med 1996;33:120.

Maher ER. Inherited renal cell carcinoma. Br J Urol 1996;78:542.

Maher ER, Yates JR, Ferguson-Smith MA. Statistical analysis of the two stage mutation model in von Hippel-Lindau disease, and in sporadic cerebellar haemangioblastoma and renal cell carcinoma. J Med Genet 1990;27:311.

Maher ER, Yates JRW, Harries R, et al. Clinical features and natural history of von Hippel-Lindau disease. Q J Med 1990;77:1151.

Mandriota SJ, Turner KJ, Davies DR, et al. HIF activation identifies early lesions in VHL kidneys: evidence for site specific tumor suppressor function in the nephron. Cancer Cell 2002;1:459.

McGuire BB, Fitzpatrick JM. BMI and the risk of renal cell carcinoma. Curr Opin Urol 2011;21:356.

Motzer RJ, Bander NH, Nanus DM. Renal-cell carcinoma. N Eng J Med 1996;335:865.

Nickerson ML, Jaeger E, Shi Y, et al. Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. Clin Cancer Res 2008;14:4726.

Parkin DM, Pisani P, Lopez AD, et al. At least one in seven cases of cancer is caused by smoking. Global estimates for 1985. Int J Cancer 1994;59:494.

Rodríguez-Perales S, Meléndez B, Gribble SM, et al. Cloning of a new familial t(3;8) translocation associated with conventional renal cell carcinoma reveals a 5 kb microdeletion and no gene involved in the rearrangement. Hum Mol Genet 2004;13:983.

Saldivar JC, Shibata H, Huebner K. Pathology and biology associated with the fragile FHIT gene and gene product. J Cell Biochem 2010;109:858.

Schoenfeld A, Davidowitz EJ, Burk RD. A second major native von Hippel-Lindau gene product, initiated from an internal translation start site, functions as a tumor suppressor. Proc Natl Acad Sci U S A 1998;95:8817.

Shuin T, Kondo K, Torigoe S, et al. Frequent somatic mutations and loss of heterozygosity of the von Hippel-Lindau tumor suppressor gene in primary human renal cell carcinomas. Cancer Res 1994;54:2852.

Sükösd F, Kuroda N, Beothe T, et al. Deletion of chromosome 3p14.2-p25 involving the VHL and FHIT genes in conventional renal cell carcinoma. Cancer Res 2003;63:455.

Thoenes W, Störkel St, Rumpelt HJ. Histopathology and classification of renal cell tumors (adenomas, oncocytomas, and carcinomas): the basic cytological and histopathological elements and their use for diagnostics. Path Res Pract 1985;181:125.

Toma MI, Grosser M, Herr A, et al. Loss of heterozygosity and copy number abnormality in clear cell renal cell carcinoma discovered by high-density Affymetrix 10K single nucleotide polymorphism mapping array. Neoplasia 2008;10:634.

Van den Berg A, Dijkhuizen T, Draaijers TG, et al. Analysis of multiple renal cell adenomas and carcinomas suggests allelic loss at 3p21 to be a prerequisite for malignant development. Genes Chromosomes Cancer 1997;19:228.

Van den Berg A, Draaijers TG, Kok K, et al. Normal FHIT transcripts in renal cell cancer- and lung cancer-derived cell lines, including a cell line with a homozygous deletion in the FRA3B region. Genes Chromosomes Cancer 1997;19:220.

Van den Berg E, Van der Hout AH, Oosterhuis JW, et al. Cytogenetic analysis of epithelial renal-cell tumors: relationship with a new histopathological classification. Int J Cancer 1993;55:223.

Wilhelm M, Bugert P, Kenck C, et al. Terminal deletion of chromosome 3p sequences in nonpapillary renal cell carcinomas: a breakpoint cluster between loci D3S1285 and D3S1603. Cancer Res 1995;55:5383.

Young AP, Schlisio S, Minamishima YA, et al. VHL loss actuates a HIF-independent senescence programme mediated by Rb and p400. Nat Cell Biol 2008;10:361.

Zimonjic DB, Druck T, Ohta M, et al. Positions of chromosome 3p14.2 fragile sites (FRA3B) within the FHIT gene. Cancer Res 1997;57:1166.

HISTONE METHYLTRANSFERASE GENE *SETD2* IS A NOVEL TUMOR SUPPRESSOR GENE IN CLEAR CELL RENAL CELL CARCINOMA

2

Gerben Duns[1], Eva van den Berg[1], Inge van Duivenbode[1], Jan Osinga[1], Harry Hollema[2], Robert M.W. Hofstra[1], and Klaas Kok[1]

Departments of [1]Genetics and [2]Pathology & Medical Biology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

Adapted from Duns et al., Cancer Research 2010;70(1):4287-91.

ABSTRACT

Sporadic clear cell renal cell carcinoma (cRCC) is genetically characterized by the recurrent loss of the short arm of chromosome 3, with a hotspot for copy number loss in the 3p21 region. We applied a method called "gene identification by nonsense-mediated mRNA decay inhibition" to a panel of 10 cRCC cell lines with 3p21 copy number loss to identify biallelic inactivated genes located at 3p21. This revealed inactivation of the histone methyltransferase gene *SETD2*, located on 3p21.31, as a common event in cRCC cells. *SETD2* is nonredundantly responsible for trimethylation of the histone mark H3K36. Consistent with this function, we observed loss or a decrease of H3K36me3 in 7 out of the 10 cRCC cell lines. Identification of missense mutations in 2 out of 10 primary cRCC tumor samples added support to the involvement of loss of *SETD2* function in the development of cRCC tumors.

INTRODUCTION

Clear cell renal cell carcinoma (cRCC) is the most prevalent subtype of renal cell carcinoma and accounts for 3% of all human malignancies (1). Originating from the epithelium of the proximal tubulus, it is genetically characterized by recurrent loss of 3p regions. These losses are thought to represent one of the two hits required to functionally inactivate tumor suppressor genes (TSG) located in these regions (2). The *Von Hippel-Lindau* gene, located at 3p25, has been identified as a TSG involved in the pathology of cRCC, and is inactivated in the majority (57–80%) of cRCCs (3, 4). Although the frequent copy number loss of 3p21 (5) also suggests the presence of one or more additional TSGs in this segment of 3p, decades of research have not led to the identification of a TSG located in this genomic segment. In an attempt to identify a putative 3p21 TSG, we applied a modified version of the "gene identification by nonsense-mediated mRNA decay inhibition" (GINI) method (6) to a panel of 10 cRCC cell lines showing copy number loss of 3p21. The GINI method allows for genome-wide identification of transcripts harboring premature termination codons (PTC). Inhibition of the nonsense-mediated mRNA decay pathway *in vitro* results in the accumulation of PTC transcripts, which can be detected using gene expression microarrays. Several groups have successfully applied modified versions of the GINI method in their search for TSGs (7–9). Using GINI, we identified the histone methyltransferase gene *SETD2/HYPB*, located on 3p21.31, as a new TSG involved in the development of cRCC.

MATERIALS AND METHODS

Cell lines

cRCC cell lines RCC-AB, RCC-ER, RCC-FG2, RCC-HS, RCC-JF, RCC-JW, RCC-MF, and RCC-WK were obtained from Cell Line Services, Eppenheim, Germany. The cells, which tested negatively for *Mycoplasma*, bacteria, and fungi, had been passed 10 to 13 times upon arrival (as stated by Cell Line Services). Experiments were performed on cells that had undergone six to eight additional passages. cRCC cell lines RCC-1, RCC-4, and RCC-5 were established by Dr. C.D. Gerharz (Institute of Pathology, University Hospital, Düsseldorf, Germany) and karyotyped. To ensure their authenticity, they were recently re-karyotyped. All cRCC cell lines were maintained in RPMI 1640 (Sigma). Embryonal kidney cell line HEK293T was obtained from American Type Culture Collection (CRL-1573) and maintained in DMEM (Sigma). MN160 and PTEC are primary cultures derived from proximal tubular epithelial cells. DNA was purified using a phenol/chloroform extraction method.

Primary tumor samples

Primary cRCC tumor samples (for details, see Supplementary Table S1) were snap-frozen in liquid nitrogen, sections were cut and the histology was reviewed by a pathologist (H. Hollema). Tissue biopsies containing at least 70% tumor cells were selected for analysis. DNA was isolated using the QIAamp DNA Micro Kit (Qiagen).

DNA samples from formalin-fixed matching normal kidney samples were obtained using a paraffin material DNA isolation protocol.

ArrayCGH

ArrayCGH was carried out using Agilent Human Genome CGH Microarray 244a slides (p/n G4411B) following the protocols of the manufacturer. Slides were scanned using the G25052C DNA microarray scanner (Agilent Technologies). Array images were analyzed using Agilent feature extraction software (v10.5.1.1) and DNA analytics (Agilent).

GINI

Cells were grown to 70% confluence and treated with emetine dihydrochloride hydrate (100 µg/mL medium) or caffeine (10 mmol/L in medium), or not treated. Following 8 hours of incubation at 37°C, cells were harvested. Experiments were performed in duplicate. RNA was isolated using the RNeasy Mini Kit (Qiagen). Transcript abundance in treated and untreated cells was measured using Human HT-12_V3_Beadchips (48K; Illumina). The Lumi package implemented in R (10, 11) was used for variance stabilizing transformation followed by robust spline normalization of the microarray data (GEO accession no. GSE20491). GeneSpring GX 10.0 (Agilent) was used to calculate treatment-induced fold changes. Putative PTC-containing transcripts were selected according to the following criteria:

1. Transcripts should show emetine- as well as caffeine-induced accumulation in one or more cell lines. Transcripts were considered to be accumulated when present in the list of 1,000 transcripts with the highest treatment-induced fold changes.
2. Transcripts showing accumulation in the majority of cell lines were considered to be so-called stress-response genes and were discarded from further analyses.
3. In untreated cells, PTC transcripts should show a significantly lower abundance than their non-PTC harboring counterparts.
4. Emetine- and caffeine-induced accumulation pushes the abundance towards a level similar to the abundance in PTEC and MN160.

Mutation analysis

Thirty-seven PCR primer sets were designed to amplify and sequence all coding sequences and exon-intron boundaries of *SETD2* annotated in the human genome (National Center for Biotechnology Information). Amplified PCR products were purified using EXOSAP IT (GE Healthcare) and sequenced using BigDye Terminator chemistry (Applied Biosystems).

Reaction products were run on the ABI3130XL Genetic Analyzer (Applied Biosystems). Sequences wereevaluated using Mutation Surveyor (SoftGenetics LCC). Cancer samples showing putative mutations were re-sequenced to eliminate any potential sequencing artifacts. To check for aberrant transcripts, cDNA fragments, obtained by reverse transcription of RNA from the cell lines, were amplified using

stepdown PCR (60°C–55°C) and visualized on a 2% agarose gel. Primer sequences are available upon request.

Western blot analysis

Cells were lysed in Triton extraction buffer [0.5% Triton X-100 (v/v), 2 mmol/L phenylmethylsulfonyl fluoride, and 0.02% NaN3 (m/v)]. Histones were isolated by acid extraction O/N in 0.2 N HCl, separated using SDS-PAGE (15%) and transferred to a nitrocellulose membrane. The following primary and secondary antibodies were used for Western analysis: rabbit anti-histone 3 (ab1791, 1:4,000; Abcam), rabbit antitrimethyl histone H3 (Lys36; ab9050, 1:750; Abcam), rabbit anti–dimethyl-histone H3 (Lys 36; 07-369, 1:2,000; Millipore) and horseradish peroxidase–>conjugated secondary antibody (goat anti-rabbit; Bio-Rad). Positive staining was visualized using ECL solution mixture from Roche. Attached antibodies were stripped from the membrane using Re-Blot Plus Strong Solution (Chemicon/Millipore) to allow reuse of the membrane.

RESULTS AND DISCUSSION

ArrayCGH analysis

In our search for candidate TSGs located at 3p21, we decided to include only cRCC-derived cell lines and primary tumor samples that showed copy number loss for 3p21. This selection was based on our arrayCGH data. Seven out of 11 cRCC cell lines and 6 out of 11 cRCC primary tumor samples lost the entire short arm of chromosome 3 (3p). Three cell lines and three primary tumor samples had copy number loss of a part of 3p, with a segment of overlap from 21.4 to 69.6 Mb from pter, including 3p21 but not the *VHL* gene, located at 3p26-p25 (Supplementary Fig. S1). One cell line (RCC-WK) did not show copy number loss of 3p21 and was not considered in the subsequent GINI analysis.

GINI analysis

To identify transcripts harboring PTCs, we performed GINI on 10 cRCC cell lines with copy number loss of 3p21. Gene expression microarrays were used to identify transcripts that were accumulated as a result of nonsense-mediated mRNA decay inhibition. Transcripts were selected for further analysis based on criteria outlined in Materials and Methods. In this study, we restricted the search for biallelic inactivated transcripts to the 3p21 region. One gene fitting all our criteria was the histone methyltransferase gene *SETD2*, located at 3p21.31. Based on the observed "transcript level profiles", we expected SETD2 to harbor PTCs in RCC-AB and RCC-ER (Fig. 1).

SETD2 mutation analysis

The presence of PTC-introducing mutations in these two cell lines was confirmed by sequence analysis of all exons and exon-intron boundaries of *SETD2*. A 1-bp deletion

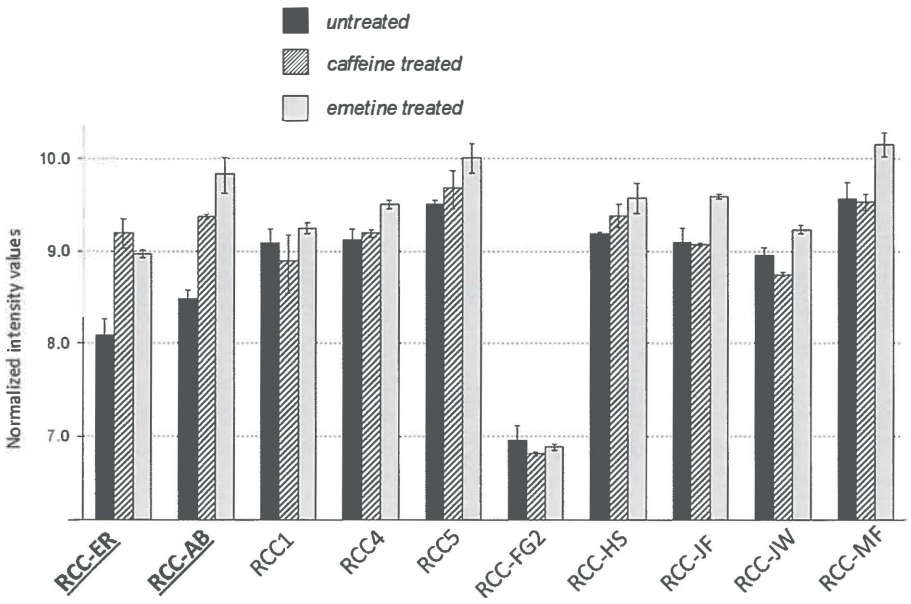


Figure 1. Effect of nonsense-mediated mRNA decay inhibition on *SETD2* transcript levels in cRCC cell lines. For each cell line, *SETD2* transcript levels are shown in untreated (black columns), caffeine-treated (striped columns), and emetine-treated (gray columns) cells, respectively. Normalized intensity values are shown on a log 2 scale. Levels below a value of 7 are close to the detection limit. The *SETD2* probe binds to the 3′ part of the transcript. The *SETD2* transcript in RCC-FG2 lacks this part (Supplementary Fig. S2.g), explaining the low transcript levels. SETD2 transcript abundance is low in untreated RCC-AB and RCC-ER cells compared with the other cell lines and shows caffeine- and emetine-induced accumulation in RCC-AB and RCC-ER. Based on these transcript level profiles, we expected RCC-AB and RCC-ER to harbor PTCs in the SETD2 gene.

resulting in a PTC 42 nucleotides downstream was detected in RCC-AB. In RCC-ER, a nonsense mutation was identified (Fig. 2; Supplementary Fig. S2; Supplementary Table S2). The other cell lines were sequenced for mutations other than those causing PTCs. We identified an in-frame 9-bp deletion in exon 1 in RCC-MF, a missense mutation in exon 3 in RCC-JW not corresponding to a known single-nucleotide polymorphism (not present in dbSNP), and a silent mutation in exon 3 in RCC-JF. In addition, the *SETD2* transcript of RCC-FG2 seemed to lack the terminal three exons (Supplementary Fig. S2). To further assess the clinical relevance of these findings, we sequenced SETD2 in 10 cRCC primary tumors with 3p21 copy number loss (Supplementary Fig. S2), and in two of them, we identified missense mutations that did not correspond to known single-nucleotide polymorphisms (dbSNP). The mutations identified in RCC-AB, RCC-ER, and RCC-MF as well as the aberrant splicing observed in RCC-FG2 will all have a severe effect on the function of the protein. The three missense mutations identified affect amino acids that are conserved across SETD2 orthologues in several species (Supplementary Fig. S3). The missense mutation in T-3 affects the SET domain, which is responsible for the methyltransferase activity of the protein. The missense

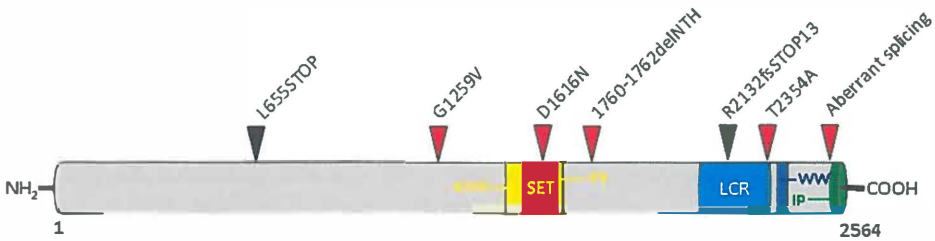


Figure 2. SETD2 mutations in cRCC samples. Schematic representation of SETD2 with arrows showing the locations of the mutations and the corresponding amino acid changes. AWS, AWS domain; SET, SET domain; PS, post-SET domain; LCR, low-charge region; WW, WW domain; IP, interaction with RNAPOL2A; Del, deletion. National Center for Biotechnology Information Reference Sequence: NP_054878.5.

mutation in T-4 affects the low charge region, which is thought to have transcriptional activation activity, whereas the missense mutation in RCC-JW affects an amino acid which is not located in a known functional domain (Fig. 2). The missense mutation in T-3 was shown to be somatic (Supplementary Fig. S2). The matching normal renal sample of T-4 was heterozygous for the missense mutation, indicating that *SETD2* mutations might contribute to genetic susceptibility for cRCC. The missense mutations in T-4 and RCC-JW were not seen in 100 healthy control samples. AGVGD, SIFT, and Polyphen gave ambiguous results in their predictions of the pathogenicity of the mutations (Supplementary Table S3). However, the ratio between observed nonsynonymous and synonymous mutations (6:1) was higher than the 2:1 predicted for nonselected passenger mutations (12), indicating that the observed mutations are functional rather than coincidental. The observed mutation frequency in the cRCC primary tumor samples might be an underestimation of the actual frequency, as the samples were not checked for aberrant splicing caused by intronic mutations.

Our findings were confirmed by a massive sequencing project on cRCC performed recently that found somatic truncating mutations in *SETD2* in 12 out of 407 (3%) cRCC clinical specimens (13). Sequencing of a panel of nine small cell lung carcinoma–derived cell lines, a type of carcinoma also characterized by loss of 3p21 (14), did not reveal any hemizygous mutations (data not shown). This implies that the inactivating mutations of *SETD2* could be specific for cRCC, as is also suggested by Dalgliesh and colleagues (13).

Trimethylation of H3K36 in cRCC

SETD2 was discovered in 1998 and initially characterized as Huntingtin-interacting protein (HYPB; refs. 15, 16). In 2005, HYPB was shown to function as a histone methyltransferase via a conserved SET domain and was renamed SETD2 ([Su(var)3–9, enhancer of zeste [E(z)] and trithorax (trx)] domain–containing protein 2). A number of recent articles correlate loss- or gain-of-function of histone-modifying enzymes, including histone lysine methyltransferases, with the pathology of several types of cancer. Aberrant activity of histone-modifying enzymes could result in altered

chromatin configuration and disruption of normal transcriptional programs, pushing the cell towards cancerous development (17). Recently, a trend for lower *SETD2* transcript levels in breast cancer tissue was shown, linking the reduction of *SETD2* mRNA levels with tumor development (18). SETD2 is nonredundantly responsible for all trimethylation of lysine 36 of histone H3 (19, 20). This mark may be involved in transcriptional elongation and splicing (21). To assess the effect of the mutations identified in *SETD2* on the global level of trimethylated H3K36, we performed Western blot analysis on our panel of cRCC cell lines. There was a strong decrease of global H3K36me3 levels in 7 out of 10 tested cRCC cell lines, whereas H3K36me2 levels were consistent across the panel of cell lines (Fig. 3). In RCC-ER and RCC-AB, which harbor hemizygous mutations introducing PTCs in SETD2, H3K36me3 was lost. The observed faint residual staining was probably an effect of nonspecific binding of the antibody. RCC-MF and RCC-FG2 have trimethylated H3K36 levels similar to those observed in RCC-AB and RCC-ER, indicating that the mutations in these cell lines totally disrupt the H3K36me3 methyltransferase activity of the SETD2 protein as well. The missense mutation identified in RCC-JW seems to have no effect on the H3K36me3 activity of the SETD2 protein, as the level of global trimethylated H3K36 was not decreased in this cell line. Given the decrease of global trimethylated H3K36 levels in a large subset of our cRCC cell line panel, global H3K36me3 might play a tumorsuppressive role in proximal epithelial tubular cells. The observed decrease of global trimethylated H3K36 in cRCC cell lines in which no mutations in *SETD2* were identified implies that, besides the loss of methyltransferase activity of SETD2, other mechanisms might also result in the loss of H3K36me3. Altogether, our results provide a strong basis for the involvement of loss of SETD2 function in the development of cRCC.

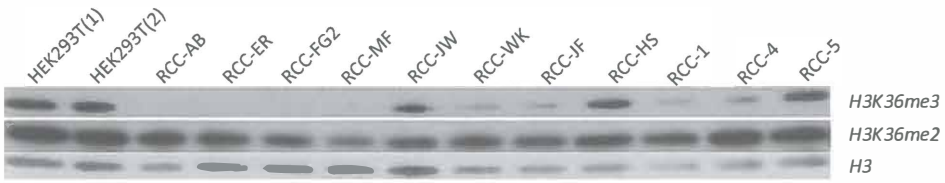


Figure 3. Global histone methylation levels in cRCC cell lines. Global H3K36me3 and H3K36me2 levels in cRCC-derived cell lines as detected by Western blot analysis. HEK293T samples were included as positive controls. Histone 3 (H3) levels were used as a loading control.

REFERENCES

Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin 2009;59:225–49.

Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971; 68:820–3.

Gnarra JR, Tory K, Weng Y, et al. Mutations of the VHL tumor suppressor gene in renal carcinoma. Nat Genet 1994;7:85–90.

Nickerson ML, Jaeger E, Shi Y, et al. Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. Clin Cancer Res 2008;14:4726–34.

Van den Berg A, Dijkhuizen T, Draaijers TG, et al. Analysis of multiple renal cell adenomas and carcinomas suggests allelic loss at 3p21 to be a prerequisite for malignant development. Genes Chromosomes Cancer 1997;19:59–76.

Noensie EN, Dietz HC. A strategy for disease gene identification through nonsense mediated mRNA decay inhibition. Nat Biotechnol 2001;19:434–9.

Ionov Y, Nowak N, Perucho M, Markowitz S, Cowell JK. Manipulation of nonsense decay identifies gene mutations in colon cancer cells with microsatellite instability. Oncogene 2004;23:639–45.

Huusko P, Poncinano-Jackson D, Wolf M, et al. Nonsense-mediated decay microarray analysis identifies mutation of EPHB2 in human prostate cancer. Nat Genet 2004;36:979–83.

Muggerud AA, Edgren H, Wolf M, et al. Data integration from two microarray platforms identifies bi-allelic inactivation of RIC8A in a breast cancer cell line. BMC Med Genet 2009;2:26–34.

Du P, Jibbe WA, Lin SM. Lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;59:1547–8.

Lin SM, Du P, Huber W, Kibbe WA. Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic Acids Res 2008; 36:e11.

Bardelli A, Parsons D, Silliman N, et al. Mutational analysis of the tyrosine kinome in colorectal cancers. Science 2003;300:949.

Dalgliesh GL, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying enzymes. Nature 2010;463:360–3.

Kok K, Osinga J, Carritt B, et al. Deletion of a DNA sequence at the chromosomal region 3p21 in all major types of lung cancer. Nature 1987;330:578–81.

Mao M, Fu G, Wu J-S, et al. Identification of genes expressed in human CD34+ hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. Proc Natl Acad Sci U S A 1998;95:8175–80.

Rega S, Stiewe T, Chang D-I, et al. Identification of the full-length huntingtin-interacting protein p231HBP/HYPB as a DNA binding factor. Moll Cell Neurosci 2001;18:68–79.

Miremadi A, Oestergaard MZ, Pharoah PDP, Caldas C. Cancer genetics of epigenetic genes. Hum Mol Genet 2007;16:Spec No. 1:R28–49.

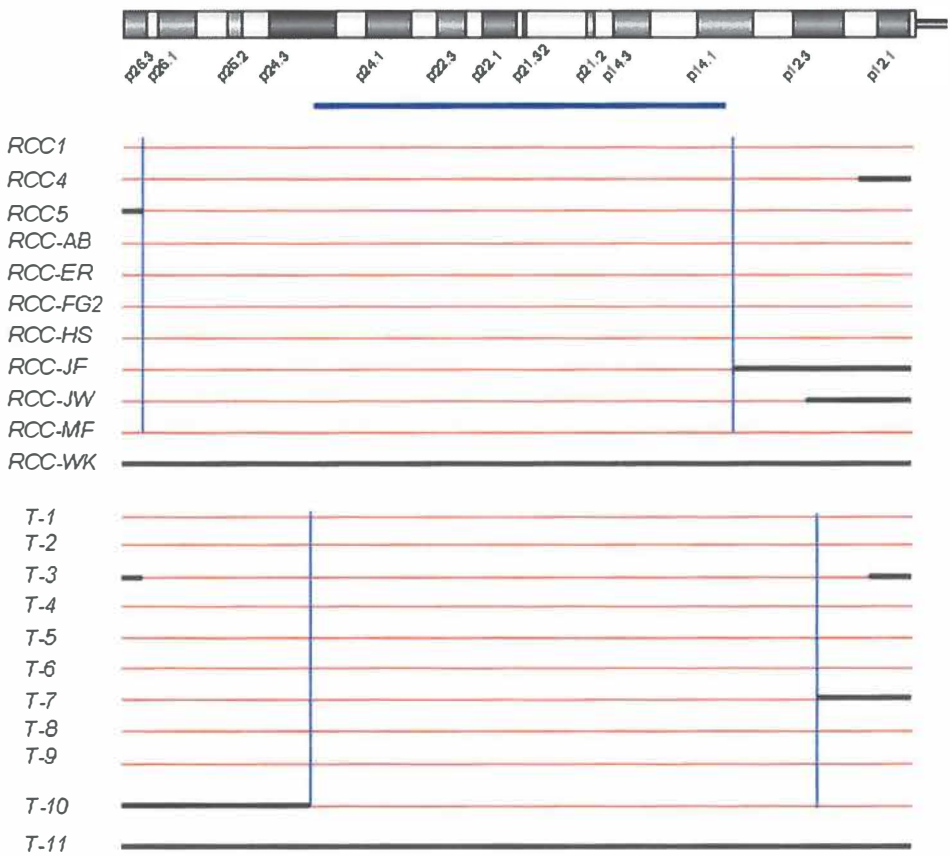
El Sarakbi W, Sasi W, Jiang WG, Roberts T, Newbold RF, Mokbel K. The mRNA expression of SETD2 in human breast cancer: correlation with clinico-pathological parameters. BMC Cancer 2009;9:290–6.

Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/SETD mediates all H3K36 trimethylation. EMBO J 2008;27:406–20.

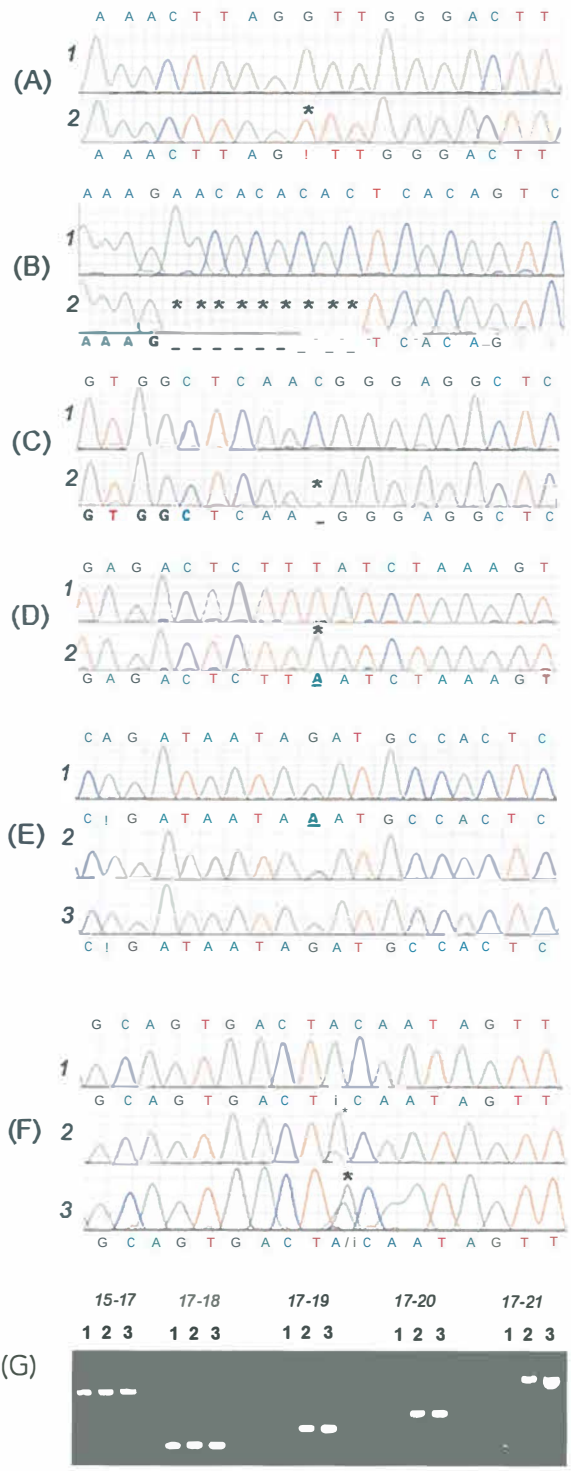
Yoh SM, Lucas JS, Jones KA. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYP/Setd2-mediated histone H3K36 methylation. Gene Dev 2008;22:3422–34.

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet 2009;41:376–81.

SUPPLEMENTARY FIGURES AND TABLES



Supplementary Figure 1. Schematic overview copy number loss of the short arm of chromosome 3 for the individual clear cell renal cell carcinoma samples. Red lines correspond to regions with copy number loss. Black lines correspond to regions which do not show copy number loss. Blue horizontal line: smallest region of overlap of copy number loss, which ranges from 21.4 Mb to 69.6 Mb from Pter.



Supplementary Figure 2. Mutation analysis *SETD2* gene: sequencing (a-f) and PCR amplification of cDNA (g) to check for mutations and aberrant transcripts. **a.** Sequence analysis of part of exon 3 of *SETD2*. (1) Wild type. (2) RCC-JW, showing the hemizygous *c.3829G>T* mutation (*). **b.** Sequence analysis of part of exon 11 of *SETD2*. (1) Wild type. (2) RCC-MF, showing the hemizygous *c.5330_5338delAACACACAC* mutation (*). **c.** Sequence analysis of part of exon 15 of *SETD2*. (1) Wild type. (2) RCC-AB, showing the hemizygous *c.6447delC* mutation (*). **d.** Sequence analysis of part of exon 3 of *SETD2*. (1) Wild type. (2) RCC-ER, showing the hemizygous *c.2017T>A* mutation (*). **e.** Sequence analysis of part of exon 7 of *SETD2*. (1) Wild type. (2) Primary clear cell renal cell carcinoma tumour *T-3*, showing the hemizygous *c.4899G>A* mutation (*). (3) Corresponding normal kidney tissue does not show the mutation, indicating that the mutation is somatic. **f.** Sequence analysis of part of exon 16 of *SETD2*. (1) Wild type. (2) Primary clear cell renal cell carcinoma tumour *T-4*, showing the hemizygous *c.7113A>G* mutation (*). (3) The same mutation was found in the corresponding normal kidney tissue, indicating a germline origin. **g.** PCR amplified cDNA fragments of exons 15-21 of the *SETD2* transcript. Cell line **1**: *RCC-FG2*, **2**: *RCC-5*, **3**: *HK-2* (*immortalized PTEC cell line*). *RCC-FG2* shows a PCR product of normal length for exons 15-17 and exons 17-18, but no product if the reversed primer is situated downstream of exon 18. This indicates aberrant splicing leading to loss of exons 19-21.

NP_054878.5	1599	NKNIHYYFMALKNDEIDATQKGNCSRFMNHSCEPNCETQKWTVNGQLRV	1648
XP_516423.2	1584	NKNIHYYFMALKNDEIDATQKGNCSRFMNHSCEPNCETQKWTVNGQLRV	1633
XP_864158.1	1597	NKNIHYYFMALKNDEIDATQKGNCSRFMNHSCEPNCETQKWTVNGQLRV	1646
XP_589886.3	1582	NKNIHYYFMALKNDEIDATQKGNCSRFMNHSCEPNCETQKWTVNGQLRV	1631
NP_001074809.1	1070	NKNIHYYFMALKNDEIDATQKGNCSRFMNHSCEPNCETQKWTVNGQLRV	1119
XP_236648.4	1460	NKNIHYYFMALKNDEIDATQKGNCSRFMNHSCEPNCETQKWTVNGQLRV	1509
XP_418510.2	1397	NKNIHYYFMALKNDEIDATQKGNCSRFMNHSCEPNCETQKWTVNGQLRV	1446

cRCC primary tumour: **T-3**. Amino acid change: *D1616N*

NP_054878.5	2345	VVQPAAAVTTIVAPGQPQPLQPSEMVVTNNLLDLPPPSPPKPKTIVLPPN	2394
XP_516423.2	2330	VVQPAAAVTTIVAPGQPQPLQPSEMVVTNNLLDLPPPSPPKPKTIVLPPN	2379
XP_864158.1	2343	VVQPATAVTTIVAPGQPQPLQPPEMVVTNNLLDLPPPSPPKPKTIVLPPN	2392
XP_589886.3	2328	VVQPTAAVTTIVAPGQPQPLQPPEMVVTNNLLDLPPPSPPKPKTIVLPPN	2377
NP_001074809.1	1815	VVQPTAAVTTIVAPGQPQSLQPPEMVVTNNLLDLPPPSPPKPKTIVLPPN	1864
XP_236648.4	2205	VVQPAAAVTTIVAPGQPQPLQPPEMVVTNNLLDLPPPSPPKPKTIVLPPN	2254
XP_418510.2	2135	IVQPPPAVTTIVAAGQPQPIQQPELVVTNNLLDLPPPSPPKPKTIVLPPN	2184

cRCC primary tumour: **T-4**. Amino acid change: *T2354A*

NP_054878.5	1253	EELRNLGWDFS-QEKPSTTYQQPDSSYGACGGHKYQQNAEQYGGTRDYWQ	1301
XP_516423.2	1238	EELRNLGWDFS-QEKPSTTYQQPDSSYGACGGHKYQQNAEQYGGTRDYWQ	1286
XP_864158.1	1249	EELRNLGWEFSQQEKPTTTYQQPDSSYGACGGHKYQQSAEHYGGTRNYWQ	1298
XP_589886.3	1234	EELRNLGWDFSQQEKPTTTYQQPDSSYGACGGHKYQQSAEQCSGARSHWQ	1283
NP_001074809.1	723	EELRNLGWDFSQQERPTTTYQQPDSSYGTCGTHKYQQSTEHYGGTHNYWQ	772
XP_236648.4	1117	EELRNLGWDFSQQERPTTTYQQPDSSYGTCGTHKYQQSAEHYGGTHDYWQ	1166
XP_418510.2	1052	EGMGGVSWELQ-AEKPSSTYQQPDSSPGVYSGYVYQPGTGAYSSSQNCWQ	1100

cRCC cell line: **RCC-JW**. Amino acid change: *G1259V*

ProteinAcc.	Gene	Organism
NP_054878.5	SETD2	Homo sapiens
XP_516423.2	SETD2	Pan troglodytes
XP_864158.1	SETD2	Canis lupus familiaris
XP_589886.3	SETD2	Bos taurus
NP_001074809.1	Setd2	Mus musculus
XP_236648.4	Setd2	Rattus norvegicus
XP_418510.2	SETD2	Gallus gallus

Supplementary Figure 3. Conservation SETD2 protein across species. Amino acids for which missense mutations were identified (in red squares) are shown together with their surrounding amino acids. Conservation is shown across 7 different species (according to *Homologene*).

Supplementary Table 1. Characteristics primary ccRCC tumor samples.

Tumour	Age at diagnosis	Sex	TNM	Stage	Diagnosis (WHO)
T-1	70	F	pT1aGxNxMx	?	Clear cell renal cell carcinoma
T-2	60	M	pT3bG3-4NxM0	III	Clear cell renal cell carcinoma
T-3	67	M	pT3bG1NxM0	III	Clear cell renal cell carcinoma
T-4	66	F	T2G1N0M0	III	Renal cell carcinoma
T-5	66	M	pT1aG2NxMx	?	Clear cell renal cell carcinoma
T-6	46	F	pT1G1NxMx	?	Clear cell renal cell carcinoma
T-7	52	M	T2G1N0M0	II	Clear cell renal cell carcinoma
T-8	84	F	pT3abG4NxMx	?	Clear cell renal cell carcinoma
T-9	68	M	pT1aG2NxM0	?	Clear cell renal cell carcinoma
T-10	82	F	pT1G2bN0Mx	?	Clear cell renal cell carcinoma
T-11	70	F	T3G3N1M0	III	Clear cell renal cell carcinoma

Supplementary Table 2. Overview of *SETD2* mutations identified in ccRCC samples

Sample	DNA mutation	Effect on **mRNA**	Effect on **protein**
ccRCC cell lines			
RCC-AB	Hem c.6447delC		R2132fsX13
RCC-ER	Hem c.2017T>A		L655X
RCC-FG2	unknown	del exons 19-21	unknown
RCC-JF	Hem.c4189T>C		none
RCC-JW	Hem c.3829G>T		G1259V
RCC-MF	Hem c.5330_5228delAACACACAC		1760-1762delNTH
ccRCC primary tumours			
T-3	c.4899G>A		D1616N
T-4	c.7113A>G		T2354A

Hem: hemizygous, del: deletion. NCBI Reference Sequence: NM_014159.6, NP_054878.5

Supplementary Table 3. Prediction program scores for identified missense mutations

	mutation		Conservation					
Sample	cDNA level	protein level	nucleotide (0.0-1.0)	Amino acid	up to	physiochemical difference	Grantham distance (0-215)	AGVGD prediction
RCC-JW	Hem c.3829G>T	G1259V	0.8	highly	Opossum	moderate	109	C15
T-3	c.4899G>A	D1616N	1.0	highly	Chicken	small	23	C15
T-4	c.7113A>G	T2354A	1.0	highly	Chicken	small	58	C55

	SIFT		PolyPhen		
Sample	predicted to	score	median sequence conservation	prediction	PSIC score difference
RCC-JW	affect protein function	0.02	4.32	possibly damaging	1.994
T-3	affect protein function	0.00	4.32	benign	1.436
T-4	affect protein function	0.00	4.32	benign	1.344

* Based on the Grantham scores, the amino acid substitution in *RCC-JW* is expected to have a significant biochemical effect.(*Grantham score:109*). The amino acid substitutions in T-3 and T-4 are not expected to have a significant biochemical effect (Grantham scores: 58 and 23).
* According to the *AGVGD* prediction program, the identified variant in *T-4* is likely to interfere with protein function (*score: C55*) , while the variants identified in T-3 and RCC-JW are less likely to interfere with protein function (score C15). (*C0: not likely to interfere with function, C65: most likely to interfere with function)*
* According to the *SIFT prediction program*, all three identified variants affect the function of the protein.(*median sequence conservation=4.32, low confidence)*
* According to the *Polyphen prediction program*, the variant identified in *RCC-JW* has a "possibly damaging"effect (*PSIC score: 1.994).* The variantes identified in T-3 and T-4 are not predicted to have a pathogenic effect (PSIC scores: 1.344 and 1.436).
Hem: hemizygous, del: deletion. NCBI Reference Sequence: NM_014159.6, NP_054878.5

EVALUATION OF THE USE OF GINI IN THE SEARCH FOR TUMOR SUPPRESSOR GENES

Duns G[1], Van den Berg E[1], Van Duivenbode I[1],
Sietzema JG[1], Osinga J[1], Hollema H[2], Hofstra RMW[1], Kok K[1]

Departments of [1]Genetics and [2]Pathology and Medical Biology,
University Medical Centre Groningen, University of Groningen,
Groningen, the Netherlands

3

ABSTRACT

Gene Identification by Nonsense mediated mRNA decay Inhibition (GINI) provides a promising possibility to identify putative tumor suppressor genes (TSGs). We applied a GINI strategy to a panel of ccRCC-derived cell lines and identified *SETD2* as a novel ccRCC tumor suppressor gene, as described in *Chapter 2*. Here, we describe the applied GINI strategy in more detail. Moreover, we evaluate the use of GINI in the search for TSGs by using the current knowledge on ccRCC genetics. We show that GINI has some intrinsic limitations. TSGs that code for relatively small proteins and/ or have only a limited number of exons appear to have a limited chance on being identified by GINI. Moreover, the succes of the method is dependent on the quality of the used gene expression platform.

INTRODUCTION

Tumor suppressor genes are in general inactivated by large deletions, mutations that alter the amino acid sequence of the protein, or mutations that introduce a premature termination codon (PTC), which include nonsense mutations, out-of-frame indels or a splice site mutation. In the majority of cases, the presence of a PTC does not result in the production of truncated proteins, but rather induces accelerated decay by nonsense mediated mRNA decay (NMD) (Anczukow et al., 2008; Karam et al., 2008; Mort et al., 2008; Perrin-Vidoz et al., 2002).

In 2001 Noensie and Dietz described a method for the genome-wide identification of transcripts that harbor a PTC in cell lines. They pharmacologically inhibited nonsense mediated mRNA decay (NMD) by the translational inhibitor emetine dihydrochlorate. This results in the accumulation of a subset of transcripts harboring PTCs and the increased transcript levels can be measured using gene expression microarrays. The presence of the PTC in the transcript can be confirmed by Sanger sequencing (Noensie and Dietz, 2001).

Gene Identification by Nonsense-mediated mRNA decay Inhibition (GINI) provides a promising possibility to identify putative tumor suppressor genes (Huusko et al., 2004; Ivanov et al., 2007). We applied a GINI strategy in our search for tumor suppressor genes involved in the development of clear cell renal cell carcinoma (ccRCC). We combined several methods and criteria to maximize the chances of finding ccRCC tumor suppressor genes and, at the same time, minimize the number of false-positive candidates that hampered the analysis of GINI strategies described earlier. Here, we describe the GINI strategy that led to our identification of histone methyltransferase SETD2 as a new ccRCC tumor suppressor (Duns et al., 2010). Moreover, using the known ccRCC tumor suppressor gene *VHL* (Gnarra et al., 1994; Nickerson et al., 2008) and the recently identified ccRCC tumor suppressor gene *PBRM1* (Chapter 5 of this thesis; Varela et al., 2011) as examples, we describe the intrinsic limitations of using GINI in the search for tumor suppressor genes.

MATERIAL AND METHODS

Cell lines

Prostate cancer cell line Du145 was obtained from the American Type Culture Collection, Manasses, Virginia, USA, and maintained in DMEM (Sigma-Aldrich, St.Louis, MO, USA). The previously described splice site acceptor mutation introducing a PTC in *MLH1* and the hemizygous nonsense mutation in *EPHB2* in DU145 (Huusko et al., 2004) were confirmed in this cell line using capillary Sanger sequencing.

The ccRCC-derived cell lines RCC-AB, RCC-ER, RCC-FG2, RCC-HS, RCC-JF, RCC-JW, RCC-MF, and RCC-WK were obtained from Cell Line Services, Eppenheim, Germany. The cells, which tested negatively for Mycoplasma, bacteria, and fungi, had been passed 10 to 13 times upon arrival (as stated by Cell Line Services). Experiments were performed on cells that had undergone six to eight additional passages. cRCC

cell lines RCC-1, RCC-4, and RCC-5 were established by Dr. C.D. Gerharz (Institute of Pathology, University Hospital, Düsseldorf, Germany) and karyotyped. To ensure their authenticity, they were recently re-karyotyped. All ccRCC cell lines were maintained in RPMI 1640 (Sigma). MN160 and PTEC are primary cultures derived from proximal tubular epithelial cells. DNA was purified using a phenol/chloroform extraction method.

ArrayCGH

ArrayCGH was carried out using Agilent Human Genome CGH Microarray 244a slides (p/n G4411B) following the manufacturer's protocols. Slides were scanned using the G25052C DNA microarray scanner (Agilent Technologies, Santa Clara, CA, USA). Array images were analyzed using Agilent feature extraction software (v10.5.1.1) and DNA analytics (Agilent).

GINI

Cells were grown to 70% confluence and either treated with emetine dihydrochloride hydrate (100 µg/mL medium) or caffeine (10 mmol/L in medium), or not treated. Following 8 hours of incubation at 37°C, cells were harvested. Experiments were performed in duplicate. RNA was isolated using the RNeasy Mini Kit (Qiagen, Venlo, the Netherlands).

Transcript abundance in treated and untreated cells was measured using Human HT-12_V3_Beadchips (48K; Illumina, San Diego, CA, USA). The Lumi package implemented in R (Du et al., 2008; Lin et al., 2008) was used for variance stabilizing transformation followed by robust spline normalization of the microarray data (GEO accession no. GSE20491). GeneSpring GX 10.0 (Agilent) was used to calculate treatment-induced fold changes.

Mutation analysis

Mutational analysis of *SETD2* in the panel of ccRCC-derived cell lines is described in *Chapter 2*. Mutational analysis of *PBRM1* and *VHL* in the panel of ccRCC-derived cell lines and primary tumors is described in *Chapter 5*. Nine and ten PCR primer sets were designed to amplify and sequence exons and exon-intron boundaries of *ARID3A*, and *ARID5B*, respectively, annotated in the human genome (National Center for Biotechnology Information). Amplified PCR products were purified using EXOSAP IT (GE Healthcare, Little Chalfont, Buckinghamshire, UK) and sequenced using BigDye Terminator chemistry (Applied Biosystems, Foster City, CA, USA). Reaction products were run on the ABI3130XL Genetic Analyzer (Applied Biosystems). Sequences were evaluated using Mutation Surveyor (SoftGenetics LCC, State College, PA, USA). The following primers were used: ARID3A: exon 1_F: CGGCCAGGTATTTGCATCCG, exon 1_R: CACCAACATCTCCGCCCAAC, exon 2a_F: CTGCATCCTGCATGAGATGG, exon 2a_R: CCATCACGGCCTGTAGTTTC, exon 2b_F:CAGGGCCATGAAACTACAG, exon 2b_R: CACTCAGAGCCAGTAACAG, exon 3_F: GCCGGCTCTTCCTATTTC-CAGAGA, exon 3_R: GCAACGTGGGCCTTCAAAGGG, exon 4_F: CTCCAGT-GCAGGAGGGACATGGT, exon 4_R: CTCACCACCGCCCCGACATACC, exon 5_F:

GGTGGATAAACCGAGGCAGAGCTG, exon 5_R: CCTCGCACCCCTACTTGACGC, exon 6_F: GGTGGTGCCACAGTGGGGTTTAC, exon 6_R: ACCCAGAGAGGGGAA-GAGGTATCC, exon 7_F: TCGCACCATTGCACTCCAG, exon 7_R: ATGTAGGCTC-CAAGGCCCTCAC, exon 8_F: CAGAGCTTGCAGTGAGCTG, exon 8_R: taccg-ggtgaacgatggtat. ARID5B:_exon 1a_F: ACATGGCAGTGAGGAAGC, exon 1a_R: AGAGAAAGGCGCCATACG, exon 1b_F: TGCCTTGATCCATCCTCTTC, exon 1b_R: ATACAGACGACCCAACAG, exon 2_F: CCCAGCGTTATATCTGTTGGGTCGT, exon 2_R: GGAGTCGATCCTGCTTGCCTGC, exon 3_F: TGGGCCTTGTGGAGAGTGGATCA, exon 3_R: ACCCATCCAGTCATCTCTCCATCC, exon 4_F: GCTGGCATCCCCCA-GACTGTAA, exon 4_R: CCCGTTCCCCATCACCGCTCTT, exon 5_F: TGTAATAAT-GAGGGCTGCCTGCCAC, exon 5_R: TGCCTCCAAGACACAGGCACAGA, exon 6_F: TACCGGTTCACCCTTAGCAC, exon 6_R: TCCAATTCAAAGAATCTGAGCA, exon 7_F: CTTTGGGCAAATTGTTTCCT, exon 7_R: CCCAATAAACCCCAAAAGT, exon 8_F: TTCTCTGTTTCTGGGTTTCCAG, exon 8_R: TTTCAGTTTCCCAAAAGGTCA, exon 9_F: CCATAGCATTGATGCTACTGGA, exon 9_R: CATCAAGCCATCCAACTGTG.

Quantitative RT-PCR

RT-PCR was performed on GINI samples to assess *PBRM1* transcript levels. RNA was reverse transcribed using the Fermentas RevertAid H minus first strand cDNA kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). SybrGreen RT-PCR was performed in a total volume of 10 ul using a 384-well microwell plates and the ABI7900HT Fast RT-PCR system (Applied Biosystems, Foster City, CA, USA) The following primers were used for *PBRM1*: Forward: TGCAGAGGCCAACCTACAAC, Reversed: TCGTGTAGCCAGGTGGAATG. The housekeeping gene *TBP* was used as a reference (forward primer: GCCCGAAACGCCGAATAT, reversed primer: CCGTGGTTCGTGGCTCTCT). All PCR experiments were performed in triplicate.

RESULTS AND DISCUSSION

NMD inhibition by emetine and caffeine treatment

Pharmacological inhibition of NMD induces a cellular stress response (Noensie and Dietz, 2001). Therefore, the list of transcripts that show an increase in abundance as a result of NMD inhibition not only includes transcripts that harbor premature termination codons ("PTC transcripts"), but also transcripts encoded by upregulated "stress response genes". Several approaches have been used in earlier GINI surveys to reduce the number of these false-positives and to obtain a limited list of candidate PTC transcripts. These approaches include a combination of NMD inhibition with transcriptional inhibition by actinomycin D (Huusko et al., 2004; Ivanov et al., 2007), and the use of siRNA-mediated knockdown of proteins essential for NMD (Huusko et al., 2004).

We anticipated that treatment of cells by compounds that block NMD via different mechanisms would excite different stress responses. A combination of treatments with different compounds could therefore be used to exclude treatment-specific stress

response genes. In an attempt to obtain a selective list of candidate PTC transcripts in ccRCC-derived cell lines, we combined two different methods of inhibiting the NMD that were used in earlier GINI screens:

(1) treating the cells with emetine dihydrochlorate (we will refer to this inhibition as emetine treatment) (Noensie et al., 2001). This inhibits the movement of the ribosome along the nascent transcripts. As a result, the so-called "pioneer round of translation", which is essential for NMD, cannot take place (Ishigaki et al., 2001).

(2) treating the cells with caffeine (Ivanov et al., 2007). This inhibits the phosphorylation of Rent1 (hUpf1) by SMG-1, a process which is essential for NMD-induced mRNA degradation (Brumbaugh et al., 2004; Ohnishi et al., 2003).

GINI validation

To validate these methods, we first applied them to the human prostate cancer cell line DU145, which harbors truncating mutations in *MLH1* and *EPHB2* (Huusko et al., 2004). Emetine- and caffeine-treatment of DU145 resulted on average in 2.6- and 2.8-fold accumulation for *MLH1* and 2.1- and 2.1-fold accumulation for *EPHB2* transcripts, respectively, as assessed by RT-PCR (*Figure 1*).

We then checked if gene expression microarrays could be used to identify these treatment-induced increases in transcript levels. *MLH1* showed emetine- and caffeine-induced fold changes of 3.9 and 2.3, respectively, for *EPHB2*, this was 1.6 and 1.4, respectively (data not shown). Thus, gene expression microarrays can be used to identify NMD-inhibition-induced increases in the abundance of PTC transcripts.

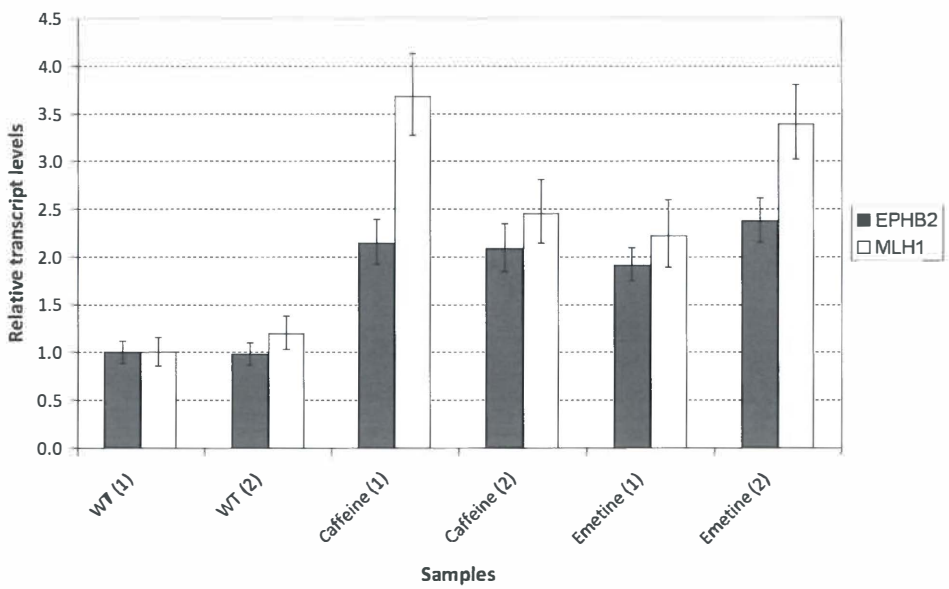


Figure 1. Increased transcript levels of *MLH1* and *EPHB2* as a result of NMD inhibition in the prostate cancer- derived cell line DU145.

GINI in ccRCC

We subsequently applied GINI in the search for tumor suppressor genes involved in ccRCC development. We used a combination of methods and criteria to increase our chances of identifying tumor suppressor genes, while at the same time limiting the number of false-positive candidate PTC transcripts.

Loss of the short arm of chromosome 3 (3p) is frequently observed in ccRCC tumors, which suggests the presence of tumor suppressor(s) in this region (Knudson et al., 1971). To increase the chance of identifying these genes, we selected ten ccRCC-derived cell lines showing copy number loss of 3p, as assessed by arrayCGH (*This thesis: Chapter 2*). We anticipated that a gene which is inactivated in a majority of tumors has a considerable chance of being inactivated by a mutation introducing a PTC that elicits NMD in at least one of the cell lines, although most likely not in all of them. In addition, we used two proximal tubular epithelial cell (PTEC) samples as reference. The transcript levels in these samples should reflect the "normal" situation. We inhibited NMD in the ccRCC-derived cell lines by treating them with emetine or caffeine.

Filtering steps

Transcript levels in treated and untreated cells were assessed using gene expression microarrays. For each ccRCC-derived cell line, lists of probes were generated that showed emetine- or caffeine-induced increase of transcript levels. We arbitrarily selected the top 1000 probes with the highest treatment-induced fold changes. To limit the lists of candidate PTC transcripts, we applied the following three criteria (summarized in Figure 2):

(1) To eliminate treatment-specific stress response genes, transcripts that showed increased transcript levels after only one of the two treatments were eliminated. This resulted in a list of 203 (range 121-321) probes on average per cell line.

(2) Genes were then selected based on their genomic location. 14 probes (representing 14 genes) that passed filtering step #1 in one or more of the cell lines were located on the short arm of chromosome 3, including the *RASSF1* gene, an isoform of which is a known physiological NMD transcript (Mendell et al., 2004).

(3) The "GINI profiles", i.e., the transcript levels in the wild-type (WT) and treated samples, of the 13 other genes were visually inspected for a so-called "PTC profile": due to NMD, PTC transcripts should be significantly less abundant than their non-PTC harboring counterparts (in PTECs as well as in the other ccRCC-derived cell lines). Emetine- and caffeine-treatment should restore the abundance of PTC transcripts to a level that resembles the wild-type transcript levels. The expected transcript levels for a hypothetical gene harboring a premature termination codon in two of the cell lines are shown in *Figure 2*.

This last filtering step, comparing transcript levels in other untreated cell lines of comparable nature as a selection criterion, has been described previously (Rossi et al., 2005). This is however clearly not without risks, as transcript levels of several genes will likely vary, even between cells of the same origin. So for genes that show fluctuating transcript levels between the different untreated samples, the criteria should be applied less stringently.

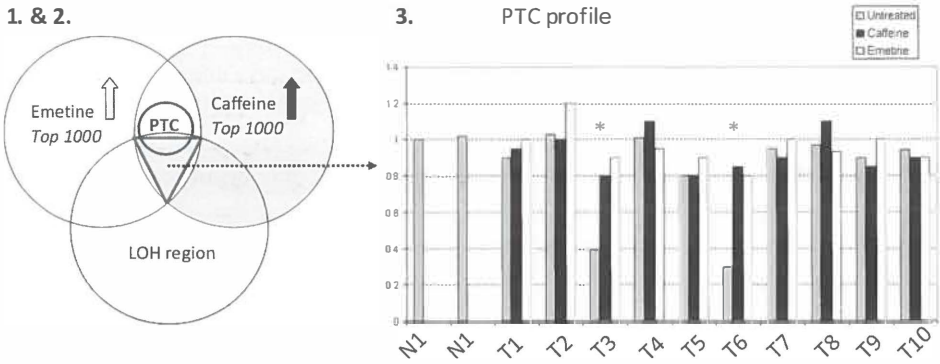


Figure 2. Overview of our GINI strategy. The graph shows the expected transcript levels for a hypothetical gene that harbors mutations introducing a premature termination codon in tumor-derived cell lines T-3 and T-6. *LOH: loss of heterozygosity, PTC: transcript harboring premature termination codons, N: "normal" cells, T: tumor-derived cell line.*

SETD2 is the only gene out of the 13 candidates on 3p that fitted all these criteria.

The GINI patterns of *SETD2* in RCC-AB and RCC-ER were indicative for the presence of a premature termination codon: in untreated RCC-AB and RCC-ER cells the *SETD2* transcript abundance was low compared to the other cell lines. Emetine- and caffeine treatment resulted in the accumulation of *SETD2* transcripts in RCC-AB and RCC-ER (*Figure 3*).

We indeed identified a 1-bp deletion introducing a frameshift and a nonsense mutation in RCC-AB and RCC-ER, respectively (Duns et al., 2010). A screen for mutations other than those causing premature termination codons in the other ccRCC-derived cell lines revealed inactivating mutations in three of them. A huge sequencing project on ccRCC tumors by Dalgliesh et al. (2010) identified truncating mutations in *SETD2* in 12 out of 407 (3%) clinical ccRCC specimens, providing *in vivo* evidence for *SETD2* being a candidate ccRCC tumor suppressor gene (Dalgliesh et al., 2010). *SETD2* is located at 3p21 and encodes a histone methyltransferase that is nonredundantly responsible for trimethylation of lysine 36 of histone 3 (H3K36me3) (Edmunds et al., 2008).

False-positives

Given the emerging role for deregulated chromatin regulation in ccRCC pathogenesis, we re-analyzed our data, and focused on PTC profiles for genes with a function in chromatin regulation. We saw possible PTC profiles for the genes coding for the chromatin remodeling proteins *ARID3A* and *ARID5B* in five and two cell lines, respectively (*Figure 4*).

However, we did not identify any truncating mutations by capillary Sanger sequencing in these genes.

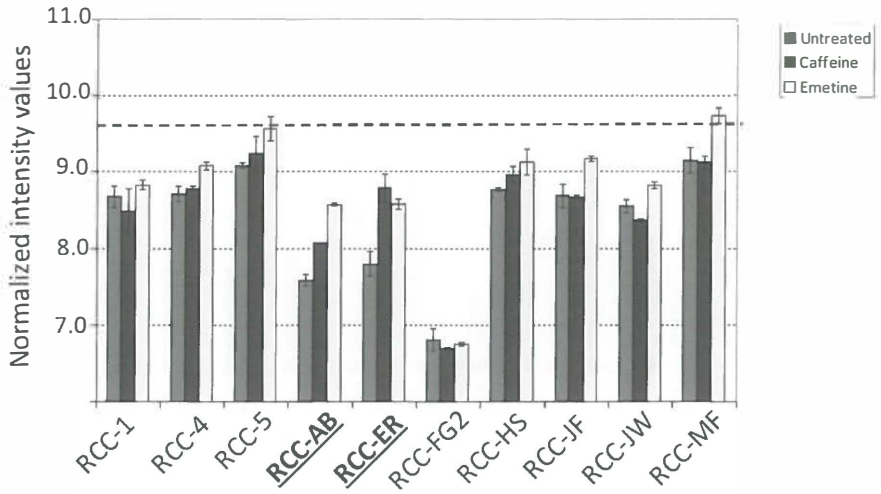


Figure 3. Effect of nonsense-mediated mRNA decay inhibition on *SETD2* transcript levels in ccRCC-derived cell lines. For each cell line, *SETD2* transcript levels are shown in untreated, caffeine-treated and emetine-treated cells, respectively. Normalized intensity values are shown on a log2 scale. Levels below a value of 7 are close to the detection limit. The dotted horizontal lines represent the average transcript levels measured in untreated proximal tubular epithelial cells.

False-negatives

With the current knowledge on ccRCC genetics and the known mutational status of *PBRM1* and *VHL* in our panel of ccRCC-derived cell lines (Chapter 5 of this thesis), we retrospectively evaluated our GINI strategy and showed that the use of GINI in searching for tumor suppressor genes has some intrinsic limitations.

The *VHL* gene, located at 3p25 and coding for the Von Hippel Lindau protein, is a major tumor suppressor gene in clear cell renal cell carcinoma. *VHL* has been shown to be inactivated in a majority of ccRCC tumors (Gnarra et al., 1994; Nickerson et al., 2008). *PBRM1*, located at 3p21, and coding for BAP180, the chromatin targeting subunit of the SWI/SNF complex, was recently identified as the second major tumor suppressor gene in ccRCC (Chapter 5 of this thesis; Varela et al., 2011). Capillary Sanger sequencing of *VHL* and *PBRM1* in our panel of ten ccRCC-derived cell lines revealed inactivating mutations in six and five of them, respectively. Three and four of the mutations identified in *VHL* and *PBRM1*, respectively, introduce premature termination codons. (This thesis: Chapter 5, table 1). However, the GINI profiles for *VHL* and *PBRM1* among the panel of ccRCC-derived cell lines do not indicate the presence of premature termination codons in any of the ccRCC-derived cell lines (*Figure 5*). Thus, *VHL* and *PBRM1* exemplify intrinsic drawbacks of using GINI in the search for tumor suppressor genes; this is explained below.

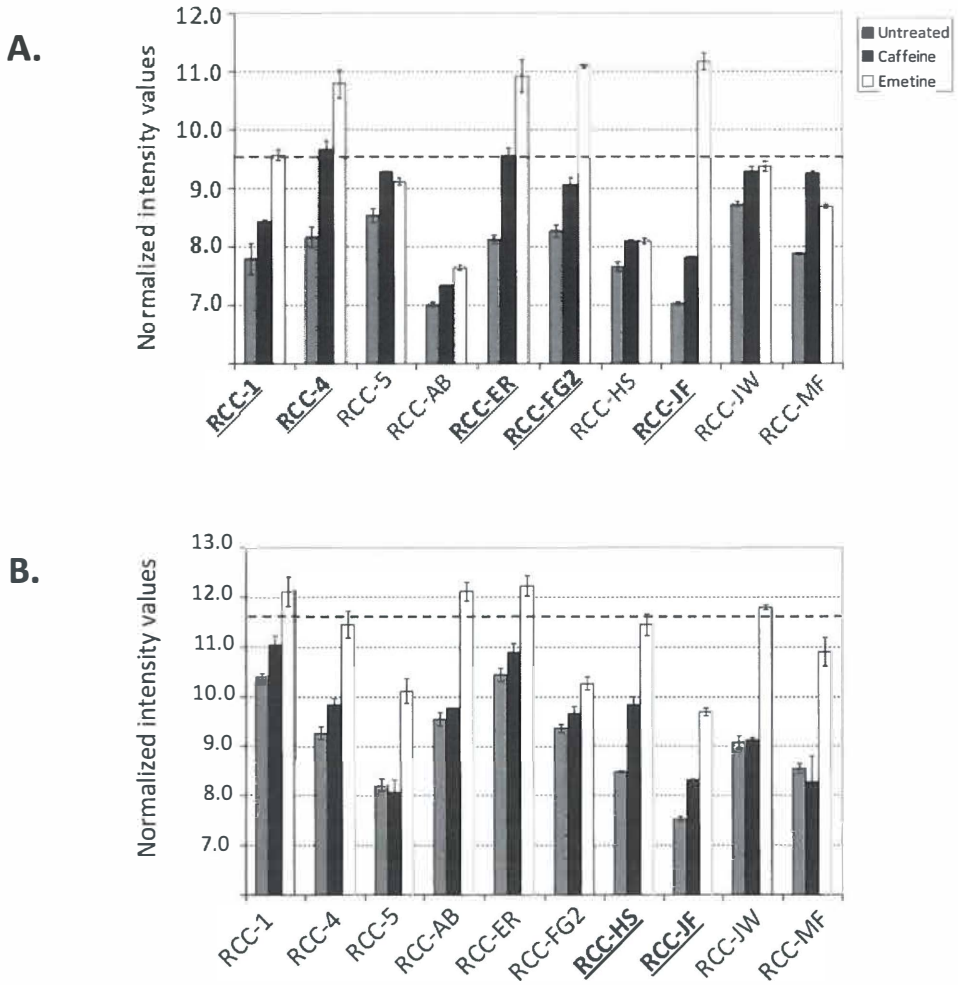


Figure 4. Effect of nonsense-mediated mRNA decay inhibition on *ARID3A* and *ARID5B* transcript levels in ccRCC-derived cell lines. For each cell line, *ARID3A* (A) and *ARID5B* (B) transcript levels are shown in untreated, caffeine-treated and emetine-treated cells, respectively. Normalized intensity values are shown on a log 2 scale. Levels below a value of 7 are close to the detection limit. The dotted horizontal lines represent the average transcript levels measured in untreated proximal tubular epithelial cells.

Size of the encoded protein and number of exons

The success of the method, i.e., the identification of a given tumor suppressor gene, depends crucially on the presence of an NMD-eliciting mutation in that gene in at least one of the tested samples. Premature termination codons that are located in the last exon, or within the last 55 nucleotides of the penultimate exon are not recognized by an NMD machinery (Nagy et al., 1998). Therefore, tumor suppressor genes that code for small proteins and/or have only a limited number of exons have a smaller chance of

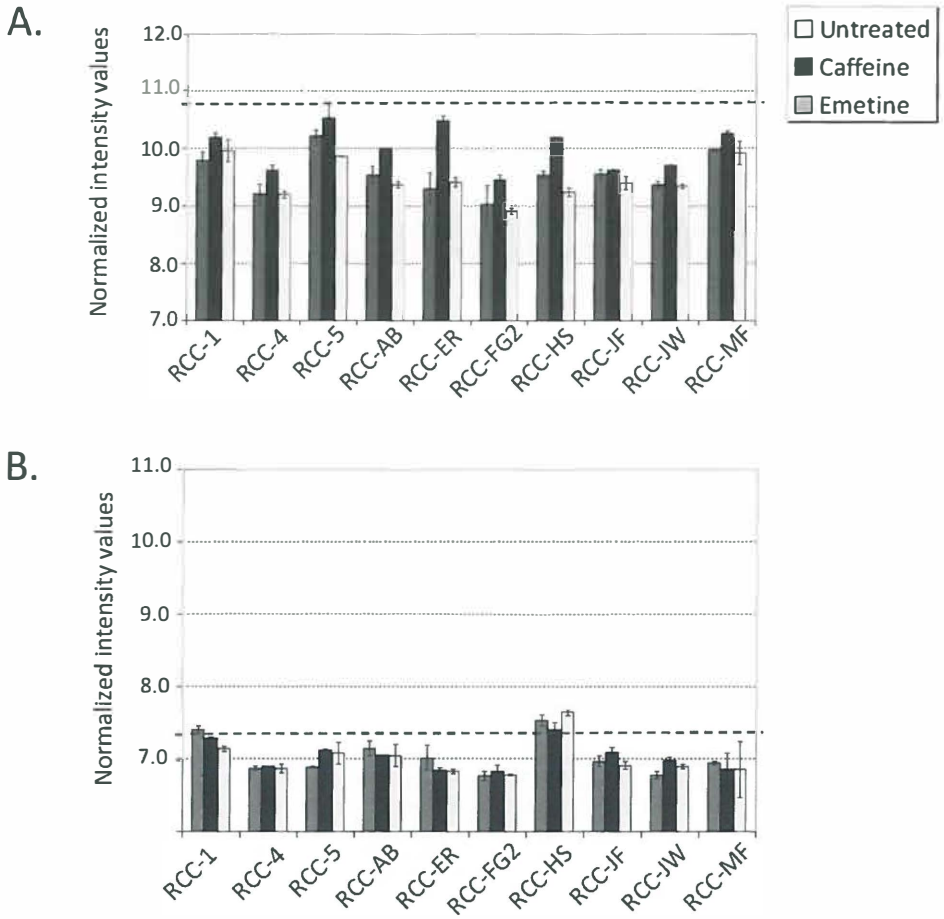


Figure 5. Effect of nonsense-mediated mRNA decay inhibition on *VHL* and *PBRM1* transcript levels in ccRCC-derived cell lines. For each cell line, *VHL* (A) and *PBRM1* (B) transcript levels are shown in untreated, caffeine-treated, and emetine-treated cells, respectively. Normalized intensity values are shown on a log2 scale. Levels below a value of 7 are close to the detection limit. The dotted horizontal lines represent the average transcript levels measured in untreated proximal tubular epithelial cells.

being inactivated by a NMD-eliciting mutation compared to genes that code for larger proteins and/or contain more exons. Even more, a PTC in a gene containing only one exon will never elicit NMD.

The full *VHL* gene contains three exons and encodes a protein (VHL_{30}) containing 213 amino acids (Iliopoulos et al., 1995). As a result, a relatively large fraction of premature termination codons in *VHL* in ccRCC tumors is expected to be in a location in which they cannot be detected by the NMD machinery (*Figure 6*.a). We identified mutations in *VHL* that introduce a premature termination codon in three out of ten ccRCC-derived cell lines. None of these are at a position where they could elicit NMD (*Figure 6*.a), which explains why *VHL* was not picked up in this GINI screen. In three

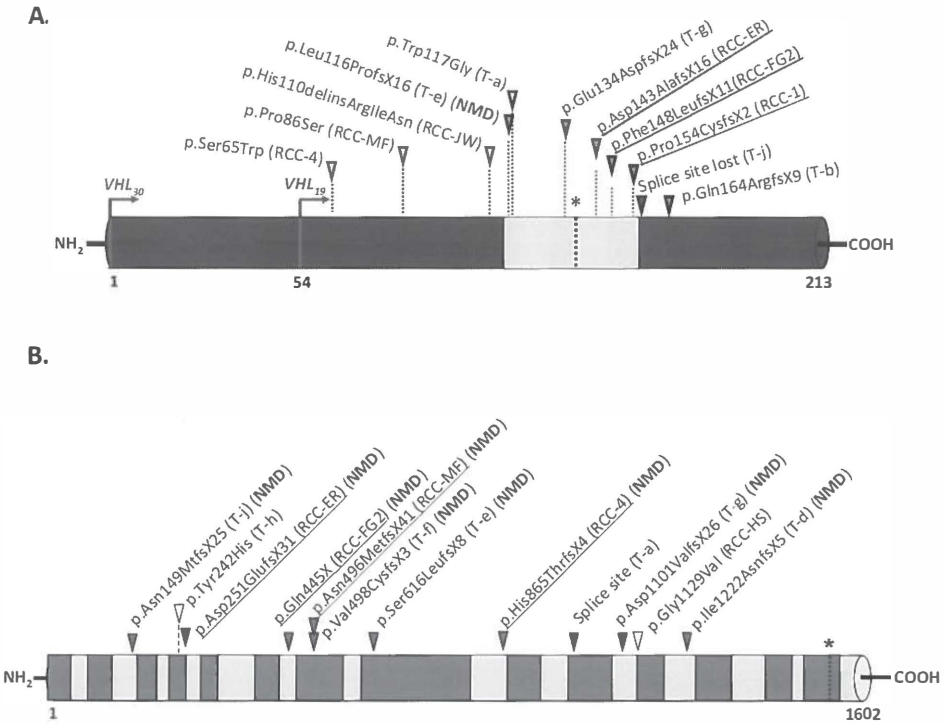


Figure 6. *VHL* and *PBRM1* mutations in ccRCC samples. Schematic overview of pVHL (A) and BAF180, encoded by *PBRM1* (B). Colors highlight the protein parts encoded by the different exons. Arrows show the location of the mutations and the corresponding amino acid changes. Underlined: mutations in ccRCC-derived cell lines that introduce a premature termination codon (PTC). PTCs located downstream of * are not recognized by the NMD machinery. For *VHL*, the region in which a PTC is not recognized by NMD includes 234 nucleotides (full *VHL* transcript: 642 nucleotides). For *PBRM1*, this region includes 228 nucleotides (full *PBRM1* transcript: 4806 nucleotides). *NMD: mutation results in a PTC that is recognized by the NMD machinery. National Center for Biotechnology Information Reference Sequence: NM_000551.3, NP_000541.2 (VHL) and NM_018165.4, NP_060635.2 (PBRM1).*

out of ten primary ccRCC tumors, we identified mutations in *VHL* that introduce a PTC. In only one primary tumor (T-e), could the PTC transcript have been potentially recognized by the NMD machinery (*Figure 6.*a). So, the chance of identifying a given tumor suppressor gene by GINI depends on the size of the encoded protein and the number of exons.

Expression platform

PBRM1 illustrates how GINI depends on the expression platform used. Screening of *PBRM1* (28 exons, 1602 amino acids) for mutations in our cell lines revealed four mutations introducing premature termination codons that are all supposed to be detected by the NMD pathway (*Figure 6.b*). However, the observed signals for *PBRM1* expression were very low in all samples, and none of the probes showed increased

transcript levels as a result of NMD inhibition in any of these four cell lines (*Figure 5.b*). Using RT-PCR on the GINI samples from RCC-4, we clearly show an increase in *PBRM1* transcript abundance as a result of both emetine- and caffeine treatment (*Figure 7*). Thus, we must conclude that the expression profiling platform we used was not sensitive enough to reliably identify the changes in transcript abundance.

Cell lines

A further limitation of the GINI strategy is its dependence on dividing cells. Therefore, tumor-derived cell lines have to be available. In order to enlarge the chance of identifying tumor suppressor genes that are mutated in only a small fraction of samples, the GINI strategy should preferably be applied to a large number of cell lines.

CONCLUSION

In conclusion, the GINI strategy has proven to be a very helpful tool in identifying putative tumor suppressor genes. However, as described above, the method has some limitations. The recent development of next-generation sequencing (NGS) technologies makes it possible to sequence large numbers of genes in parallel, making the use of GINI redundant. The method might, however, still be useful in specific cases. For example, Sumitsuji et al. (2003) showed that NMD inhibition followed by cDNA sequencing can be used to rapidly identify mutations in the *MLH1* gene, which is large and difficult to sequence, in patients suspected of having hereditary nonpolyposis colon cancer.

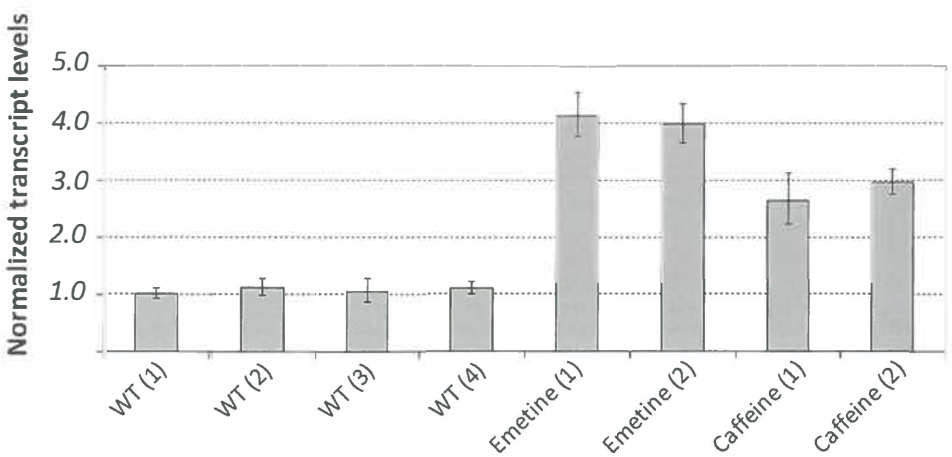


Figure 7. Effect of nonsense-mediated mRNA decay inhibition on *PBRM1* transcript levels in ccRCC-derived cell line RCC-4 as assessed by RT-PCR.

REFERENCES

Anczukow O, Ware MD, Buisson M, et al. Does the nonsense- mediated mRNA decay mechanism prevent the synthesis of truncated BRCA1, CHK2, and p53 proteins? Hum Mut 2008;29:65.

Brumbaugh KM, Otterness DM, Geisen C, et al. The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. Mol Cell 2004;14:585.

Dalgliesh GL, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying enzymes. Nature 2010;463:360.

Du P, Jibbe WA, Lin SM. Lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;59:1547.

Duns G, Van den Berg E, Van Duivenbode I, et al. Histone Methyltransferase Gene SETD2 Is a Novel Tumor Suppressor Gene in Clear Cell Renal Cell Carcinoma. Cancer Res 2010;70:4287.

Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/ SETD2 mediates all H3K36 trimethylation. EMBO J 2008;27:406.

Gnarra JR, Tory K, Weng Y, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. Nat Genet 1994;7:85.

Huusko P, Ponciano- Jackson D, Wolf M, et al. Nonsense- mediated decay microarray analysis identifies mutations of EPHB2 in human prostate cancer. Nature Genet 2004;36:979.

Iliopoulos O, Kibel A, Gray S, et al. Tumour suppression by the human von Hippel-Lindau gene product. Nat Med 1995;1:822.

Ishigaki Y, Li X, Serin G, et al. Maquat LE. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense- mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell 2001;106:607.

Ivanov I, Lo KC, Cowell JK, et al. Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense- mediated mRNA decay in colon cancer cells. Oncogene 2007;26:2873.

Kaelin Jr. WG. Von Hippel-Lindau disease. Anny Rev Pathol Mech Dis 2007;2:145.

Karam R, Carvalho J, Bruno I, et al. The NMD mRNA surveillance pathway downregulates aberrant E-cadherin transcripts in gastric cancer cells and in CDH1 mutation carriers. Oncogene 2008;27:4255.

Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68:820.

Lin SM, Du P, Huber W, Kibbe WA. Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic Acids Res 2008;36:e11.

Mendell JT, Sharifi NA, Meyers JL, et al. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet 2004;36:1073.

Mort M, Ivanov D, Cooper DN, et al. A meta-analysis of nonsense mutations causing human genetic disease. Hum Mut 2008;29:1037.

Nagy E, Maquat LE. A role for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci 1998;23:198.

Nickerson ML, Jaeger E, Shi Y, et al. Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. Clin Cancer Res 2008;14:4726.

Noensie EN, Dietz HC. A strategy for disease gene identification through nonsense- mediated mRNA decay inhibition. Nat Biotechnol 2001;19:434.

Ohnishi T, Yamashita A, Kashima I, et al. Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. Mol Cell 2003;12:1187.

Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, et al. The nonsense- mediated mRNA decay pathway triggers degradation of most *BRCA1* mRNAs bearing premature termination codons. Hum Mol Genet 2002;11:2805.

Rossi MR, Hawthorn L, Platt J, et al. Identification of inactivating mutations in the JAK1, SYNJ2, amd CPLTM1 genes in prostate cancer cells using inhibition of nonsense- mediated decay and microarray analysis. Cancer Genet Cytogenet 2005;161:97.

Sumitsuji I, Sugano K, Matsui T, et al. Frequent genomic disorganisation of MLH1 in hereditary non-polyposis colorectal cancer (HNPCC) screened by RT-PCR on puromycin treated samples. J Med Genet 2003;40:e30.

Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 2011;469:539.

THE EFFECT OF SETD2 KNOCKDOWN ON TRANSCRIPT LEVELS AND ALTERNATIVE SPLICE SITE CHOICE IN PROXIMAL TUBULAR EPITHELIAL CELLS

Duns G[1], Bilenky M[2], Sietzema JG[1], Van den Berg E[1], He A[2], Thiessen N[2], Olivier M[1], Hofstra RMW[1], Jones SJM[2], Kok K[1]

[1]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
[2]BC Cancer Agency-Genome Sciences Centre, Vancouver, British Columbia, Canada

4

ABSTRACT

SETD2 was recently identified as a new candidate ccRCC tumor suppressor gene. *SETD2* codes for a histone methyltransferase that is responsible for trimethylation of lysine 36 of histone H3 (H3K36me3). The specific role of SETD2 in ccRCC development is largely unknown. H3K36me3 is involved in alternative splice site choice, while misregulation of splicing is common in cancer and plays a role in tumor development and maintenance. One could therefore speculate that *SETD2* inactivation contributes to ccRCC development by the deregulated splicing of target genes. Here, we describe the initial steps to determine the effects of functional loss of SETD2 in proximal tubular epithelial cells and its contribution to ccRCC development. We show significantly altered transcript levels for a subset of genes as a result of siRNA-mediated knockdown: this subset is enriched for genes that function in angiogenesis-related processes. This is in concordance with the observation that SETD2 is required for embryonic vascular remodeling in mice, and might indicate a role for SETD2 in tumor angiogenesis. Moreover, several genes show an apparent change in alternative splice site choice as a result of SETD2-knockdown, which corroborates earlier findings. Further functional studies are needed to determine the exact mechanisms and proteins that underlie the observed changes in transcript levels and alternative splice site choice, and the effect of these changes on the function of the encoded proteins, as well as their potential role in ccRCC tumorigenesis.

INTRODUCTION

SET domain-containing protein 2

SET domain-containing protein 2 (*SETD2*) was recently identified as a candidate tumor suppressor gene in clear cell renal cell carcinoma (ccRCC) (Dalgliesh et al., 2010; Duns et al., 2010). *SETD2* codes for a histone methyltransferase which is nonredundantly responsible for trimethylation of lysine 36 of histone H3 (H3K36me3) (Edmunds et al., 2008; Yoh et al., 2008). We showed inactivating mutations in *SETD2* in five out of ten ccRCC-derived cell lines. These five cell lines, and two additional ccRCC-derived cell lines, all showed loss or a significant decrease of global H3K36me3. Our data support the findings of Dalgliesh et al. (2010) who identified inactivating mutations in *SETD2* in 12 out of 407 (3%) primary ccRCC tumors. Inactivating mutations in *SETD2* appear to be specific for ccRCC (Dalgliesh et al., 2010). However, decreased *SETD2* transcript levels have been observed in malignant breast cancer (Al Sarakbi et al., 2009), and a transposon-based genetic screen in mice suggested a role for *SETD2* inactivation in the development of pancreatic cancer (Starr et al., 2009).

The precise role of *SETD2* inactivation in ccRCC development is unknown, although it is known that SETD2 is required for embryonic vascular remodeling in mice (Hu et al., 2010), which might point to a role in tumor angiogenesis. Xie et al. (2008) showed that SETD2 can directly interact with P53, and selectively regulates a subset of P53 target genes. This might suggest a tumorigenic role for *SETD2* inactivation through deregulation of the P53 pathway. The identification of inactivating mutations in additional histone modifying enzymes in a subset of ccRCC tumors (Dalgliesh et al.,2010) however strongly suggests that the tumorigenic effect of *SETD2* inactivation is probably the result of the loss of its histone methyltransferase activity.

Histone modifications and cancer

Post-translational modification of histone tails can affect chromatin structure and/or result in the recruitment or rejection of specific protein complexes, affecting gene regulation. A number of recent papers correlate loss- or gain-of-function of histone-modifying enzymes, including histone lysine methyltransferases, with the pathology of several types of cancers (reviewed by Füllgrabe et al., 2010; Morin et al., 2011). Aberrant activity of histone- modifying enzymes could result in altered chromatin configuration and disruption of normal transcriptional programs, pushing the cell towards cancerous development.

Trimethylation of histone 3 at position K36

The H3K36me3 mark is mainly located on the gene body of active genes and is therefore associated with actively transcribed genes (Barski et al., 2007). Several observations suggest the involvement of H3K36me3 and other histone modifications in the regulation of splicing. H3K36me3 is specifically enriched in exons relative to intronic regions (Andersson et al., 2009; Huff et al., 2010; Kolasinska-Zwierz et al., 2009; Spies et al., 2009). Furthermore, H3K36me3 is less prominently enriched

in alternatively spliced exons than in constitutive exons (Andersson et al., 2009; Kolasinska-Zwierz et al., 2009). Moreover, H3K36me3 peaks at bona fide exons, but not at intronic exon-like sequence composition regions (ECRs), which have high nucleosome occupancy like that of bona fide exons, but are not spliced (Huff et al., 2010). Recent reports provide more direct evidence for the involvement of H3K36me3 in alternative splicing. Depolarization of human neuronal cells increases H3K36me3 methylation locally around the alternatively spliced exon 18 of *NCAM* and induces exon skipping (Schor et al., 2009). It has been suggested that histone modifications are able to modulate the outcome of alternative splicing by influencing the elongation rate of RNAPOLII (Schor et al., 2009). Slowing of the elongation rate could result in recognition of "weak" splice sites, resulting in inclusion of alternative exons (De la Mata et al., 2003). Recently, Luco et al. (2010) showed that H3K36me3 can directly affect splicing by the recruitment of polypyrimidine tract-binding protein (PTB) via chromatin-binding protein MRG15, which specifically recognizes trimethylated H3K36. PTB binds to exon-specific silencing elements, resulting in the repression of these exons (Luco et al., 2010). They suggest the existence of several of these "adaptor complexes", which consist of a combination of: (1) a modified histone tail, (2) a chromatin-binding protein that recognizes this specific histone modification, and (3) a splice regulator that is recruited to the histone mark via the chromatin-binding protein (reviewed by Luco et al., 2011).

Alternative splicing and cancer

About 90% of human pre-mRNAs can be alternatively spliced, which results in the production of protein isoforms. Expression of these isoforms is in many cases tissue-specific, dependent on the balance of splicing enhancer and repressor activities that are present. Alternative splicing plays an important role in differentiation and development-al programs and contributes to cell lineage and tissue identity (Allemand et al., 2008; Wang et al., 2008). The importance of alternative splicing is apparent from the numerous diseases that have been shown to be caused by mutations in either *cis*-acting or *trans*-acting protein splicing factors (Caceres & Kornblihtt, 2002; Cooper et al., 2009). Misregulation of splicing is common in cancer and plays a role in tumor development and maintenance (Venables et al., 2004; 2009). Recently, it was shown that disturbed expression of splicing factors in ccRCC may lead to impaired alternative splicing of genes regulating tumor growth and may contribute to the process of carcinogenesis (Piekielko-Witkowska et al., 2010).

Aim

This chapter describes the initial steps to determine the effects of functional loss of SETD2 in proximal tubular epithelial cells (PTECs) that are thought to be the progenitor cells of ccRCCs, and its contribution to ccRCC development. Here, we aim to identify the genes that carry a H3K36me3 mark in proximal tubular epithelial cells, and assess the effect of H3K36me3 knockdown on expression and alternative exon usage of target genes in these cells.

MATERIALS AND METHODS

Proximal tubular epithelial cells and cell lines

PTEC-2A09 and PTEC-060123 are primary cultures derived from proximal tubular epithelial cells (Pathology Department, University Medical Centre, Groningen, the Netherlands). PTECs were grown in renal epithelial cell basal medium supplied with renal epithelial cell growth kit (ATCC, PCS-400-030, PCS-400-040). ccRCC-derived cell line RCC-AB was obtained from Cell Line Services, Eppenheim, Germany. The cell line, which tested negatively for *Mycoplasma*, bacteria, and fungi, had been passed 10 to 13 times upon arrival (as stated by Cell Line Services). RCC-AB was maintained in RPMI1640 (Sigma, St. Louis, MO, USA). Embryonal kidney cell line HEK293T was obtained from American Type Culture Collection (CRL-1573) and maintained in DMEM (Sigma). Experiments were performed on cells that had undergone six to eight additional passages.

siRNA experiments

Proximal epithelial cells were grown to a confluence of near 100% in 25 or 75 cm^2 flasks (37°C, 5% CO_2). The cells were transfected with *SETD2* on-target plus siRNA or Non Targeting siRNA (Dharmacon, Lafayette, CO, USA) at 100 nM final concentration using lipofectamine 2000 reagent (Invitrogen, Carlsbad, CA, USA), Mock transfected, or not treated. Following 72 h incubation, cells were harvested. Pellets were split in two fractions, and RNA was isolated from one fraction using the Fermentas Genejet RNA purification kit according to the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA, USA). Decreases in SETD2 transcript levels were assessed by quantitative RT-PCR with a Taqman gene expression assay for *SETD2* (Hs01015881_m1) (Applied Biosystems, Foster City, CA, USA), and normalized against housekeeping gene *TBP* (forward primer: GCCCGAAACGCCGAATAT, reverse primer: CCGTGGTTCGTGGCTCTCT), using the ABI9700 HT Sequence detector system (Applied Biosystems). The other part of the cell pellet was used for histone modification analysis. Cells were lysed in Triton extraction buffer [0.5% Triton X-100 (v/v), 2 mmol/L phenylmethylsulfonyl fluoride, and 0.02% NaN_3 (m/v)]. Histones were isolated by acid extraction O/N in 0.2 N HCL, separated using SDS-PAGE (15%) and transferred to a nitrocellulose membrane. The following primary and secondary antibodies were used for Western analysis: rabbit anti- histone 3 (Ab1791,1:4,000;Abcam, Cambridge, UK), rabbit anti-trimethyl histone H3 (Lys36; ab9050, 1:750; Abcam) and horseradish peroxidase-conjugated secondary antibody (goat anti-rabbit: Bio-RAD, Hercules, CA, USA). Positive staining was visualized using ECL solution mixture from Roche, Basel, Switzerland. Attached antibodies were stripped from the membrane using Re-Blot Plus Strong Solution (Chemicon/Millipore) to allow reuse of the membrane.

RNA-seq sample preparation

RNA was converted into a library of template molecules suitable for subsequent cluster generation using the Illumina TruSeq™ RNA Sample preparation kit (Illumina, San Diego, CA, USA). In brief, 400 ng of total RNA was ethanol precipitated, and

the pellet was resuspended in a fragmentation mix. The RNA was subsequently fragmented on a thermal cycler (94°C for 8 minutes). Fragmented RNA was reversibly transcribed into cDNA using SuperScript II reverse transcriptase and random primers on a thermal cycler (25°C for 10 minutes, 42°C for 50 minutes, 70°C for 15 minutes), following the manufacturer's instructions. Second strand cDNA was subsequently synthesized by adding a Second strand master mix and incubating at 16°C for a hour. Double stranded cDNA was separated from the second strand master mix using Ampure XP beads (Beckman Coulter Genomics, Danvers, MA, USA), following the manufacturer's instructions. 3' and 5' overhangs that result from fragmentation were subsequently removed by adding *End repair mix* to the purified double stranded cDNA, and incubating at 30°C for 30 minutes using thermal cycler. To prevent the blunt-end fragments from forming concatamers during the adapter ligation reaction, a single 'A' nucleotide was added to the 3' ends by adding *A-Tailing mix* to the blunt-end fragments, and incubating at 37°C for 30 minutes using a thermal cycler. Multiple indexing adapter were added to the ds cDNA by adding DNA ligase mix and RNA adapter index, and subsequently incubating at 30°C for 10 minutes in a thermal cycler. Adapter-ligated ds cDNA was purified using Ampure XP beads (Beckman Coulter Genomics), following the manufacturer's instructions. DNA fragments that had adapter molecules on both ends were subsequently enriched and amplified using PCR. PCR Master mix and a PCR Primer cocktail, which anneals to the ends of the adapters, were added to the adapter-ligated ds cDNA. The adapter-ligated ds cDNA was subsequently amplified in a PCR-reaction (98°C for 30 seconds, 15 cycles of (98°C for 10 seconds, 60°C for 30 seconds, 72°C for 30 seconds), 72°C for 5 minutes) using a thermal cycler. The amplified adapter-ligated cDNA library was subsequently purified using Ampure XP beads (Beckman Coulter Genomics), following the manufacturer's instructions. Quality and quantity of the obtained adapter-ligated cDNA library was assessed using the Experion™ DNA 12K analysis kit (Bio-RAD, Hercules, CA, USA). DSN (Duplex-specific thermostable nuclease) normalization was subsequently used to deplete abundant sequences such as those derived from rRNA and tRNA following the DSN Normalization Sample Preparation Guide (Illumina, San Diego, USA). This obviates the use of Poly-A selection, which has a bias towards the 3' end of transcripts, which could be problematic when analyzing alternative splicing. In short, a reaction mix containing 100 ng of the cDNA sample prep and hybridization buffer (50 mM Hepes, pH 7.5, 0.5 M NaCL) was prepared in a PCR tube. Double stranded cDNA molecules were denatured to single stranded DNA molecules, and subsequently renaturated by incubating the tube on a thermal cycler at 98°C for 2 minutes followed by 68°C for 5 hours. Double stranded DNA, which represent the highly abundant transcripts, were subsequently removed from the reaction mixture by treatment with Specific thermostable nuclease (DSN) enzyme (Evrogen, Moscow, Russia) at 68°C, for 25 minutes. DSN treated DNA was subsequently enriched using Phusion polymerase (Finnzymes Oy, Espoo, Finland), and purified with SPRI beads (Agencourt, Beverly, MA, USA) as described in the DSN Normalization Sample Preparation Guide (Illumina). Quality and quantity of the DSN treated cDNA library was assessed using the Experion™ DNA 12 K analysis kit (Bio-Rad).

Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed using the Simple ChIP® Enzymatic Chromatin IP Kit (Magnetic Beads) # 9003 (Cell Signaling Technology, Beverly, Massachusetts, USA), according to the manufacturer's instructions. Briefly, approximately 2 x 10^7 (proximal tubular epithelial cells) to 4 x 10^7 (HEK293T) cells were used for each experiment. DNA was cross-linked to histones by treating cells with 37% formaldehyde (10 mins, RT). Cross-linking was stopped by adding glycine to a final concentration of 0.125M (5 mins, RT). Cells were subsequently washed with PBS and harvested by scraping. Cell pellets were resuspended in cell lysis buffer with protease inhibitor and incubated for 10 mins on ice. Nuclei were subsequently pelleted by centrifugation. Chromatin was subsequently fragmented so that the majority of fragments consisted of one nucleosome, using micrococcal nuclease digestion at 37°C for 20 minutes. Nuclease digestion was stopped by adding EDTA to a final concentration of 5 mM. Nuclei were pelleted and resuspended in a ChIP buffer. Nuclei were lysed using a Bandelin sonicator (3 sets of 20 second pulses at 35% amplitude) (Bandelin, Berlin, Germany). Breakdown of the nuclear membrane was determined by observing nuclei under a light microscope. A fraction of the cross-linked preparation was used for analysis of DNA digestion and determination of chromatin concentration. DNA was isolated from this fraction using spin columns, and DNA fragment size was determined by electrophoresis on a 1% agarose gel. DNA concentration was assessed using a Nanodrop 8000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Chromatin immunoprecipitation was performed on the cross-linked chromatin preparation (O/N, 4°C), according to the manufacturer's instructions . The following antibodies (all from Cell Signaling Technology), which were kindly made available to us by Bioké, Leiden, the Netherlands, were used: Tri-Methyl-Histone H3 (Lys 36) (D5A7) XP™ Rabbit mAb (1:50), Tri-Methyl-Histone H3 (Lys4) (C42D8) Rabbit mAb (1:50), Tri-Methyl-Histone H3 (Lys27) (C36B11) Rabbit mAb (1:100), Histone 3 (D2B12) XP® Rabbit mAb (ChIP Formulated) #4620 (1:50), Normal Rabbit IgG #2729 (1:100). RT-PCR was used to assess the specific enrichment as a result of IP with the histone modification-specific antibodies. The following primer sets (all from Cell Signaling Technology) were used: H3K36me3: SimpleChIP™ Human GAPDH intron 2 (#4478) vs SimpleChIP™ Human GAPDH promoter (#4471). H3K4me3: SimpleChIP™ Human GAPDH promoter (#4471) vs SimpleChIP™ Human MyoD1 Exon 1 (#4490), H3K27me3: Human MyoD1 Exon 1 (#4490) vs SimpleChIP™ Human GAPDH promoter (#4471). IP samples were incubated overnight at 4°C with rotation. CHIP Grade Protein G Magnetic Beads # 9600 were used to collect the immunocomplexes (2h, 4°C). The Protein G Magnetic beads were subsequently pelleted by placing the tubes in a Magnetic Separation Rack. After salt-washing the beads, chromatin was eluted from the antibody/Protein G beads by gently vortexing in ChIP Elution Buffer (30 mins, 65°C). Protein G Magnetic beads were pelleted by placing the tubes in a magnetic separation rack, and the eluted chromatin supernatant was transferred to new tube. 6 µl 5M NaCl and 2 µl Proteinase K was added, and incubated for 2 hours at 65°C. DNA was subsequently purified using spin columns, and quantified using a real-time quantitative PCR method.

ChIP-seq sample preparation

Chromatin immunoprecipitated DNA was prepared for next-generation sequencing (NGS) analysis following the "Preparing Samples for ChIP sequencing of DNA" manual (Illumina). 3' and 5' overhangs were converted into phosphorylated blunt ends, by incubating a reaction mix containing 30 µl ChIP enriched DNA (~10 ng), 10 µl H_2O, 4 µl T4 DNA ligase buffer with 10 mM ATP, 2 µl dNTP mix, 1 µl T4 DNA polymerase, 1 µl Klenow DNA polymerase and 1 µl T4 PNK) for 30 minutes at 20°C. End-repaired DNA was purified using the QIAquick PCR purification kit, and eluted in 34 µl Elution buffer, according to the manufacturer's instructions (Qiagen, Venlo, the Netherlands). 'A' bases were added to the 3'end, by incubating a reaction mix containing 34 µl DNA sample, 5 µl Klenow buffer, 10 µl dATP, and 1 µl Klenow exo (3' to 5' exo minus) for 30 minutes at 37°C. A-tailed DNA fragments were purified using the minElute PCR purification kit, and eluted in 10 µl Elution buffer, according to the manufacturer's instructions (Qiagen). Adapters (that have a single 'T' base overhang at their 3' end) were subsequently ligated to the A-tailed DNA fragments by incubating a reaction mix containing 10 µl DNA sample, 15 µl DNA ligase buffer, 1 µl diluted adapter oligo mix and 4 µl DNA ligase for 15 minutes at RT. The adapter- ligated DNA fragments were purified using the minElute PCR purification kit, and eluted in 10 µl Elution buffer (Qiagen). The adapter-ligated library was subsequently size-selected (range 200-300 bp) by running the DNA sample on a 2.5% agarose gel in TAE buffer. The region of the gel containing the appropriate sized DNA fragments was excised using a scalpel. DNA was purified from the agarose slices using the Qiagen Gel extraction kit, and eluted in 36 µl Elution Buffer, according to the manufacturer's instructions (Qiagen). Adapter-modified DNA fragments were subsequently amplified in a reaction mix containing 36 µl DNA, 10 µl 5x Phusion* buffer, 1.5 µl dNTP mix, 1.1 µl PCR primer 1.1, 1 µl PCR primer 2.1 and 0.5 µl Phusion polymerase using the following PCR protocol: 98°C for 30 seconds, 18 cycles of (98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds), 72°C for 5 minutes. The adapter-modified DNA fragments were subsequently purified using the minElute PCR purification kit (Qiagen).

Sequencing

Enriched bar-coded libraries were sequenced with 100 bp paired-end reads (RNA-seq) or single-end reads (ChIP-seq) across one lane of an Illumina HiSEQ 2000 flow cell (Illumina).

RNA-Seq analysis

The resulting 100 bp sequence reads were aligned to the human reference genome (hg19) and resource containing sequences spanning annotated exon-exon junctions. The Burrows-Wheeler aligner (BWA) (v0.5.7) (Li & Durbin, 2009) was used with default setting except that the Smith-Waterman alignment for the unmapped mate was disabled. Exon-exon junctions were compiled using transcript models from Ensembl v61 (Flicek et al., 2010), RefSeq (Pruitt and Maglott, 2001) and UCSC 'known genes' (Rhead et al., 2010) gene annotation sources (downloaded from http://genome.ucsc.

edu on March 17, 2009), as described (Morin et al., 2010). Sequence reads that were uniquely aligned to junction sequences were computationally repositioned to assign the genomic hg19 coordinates and an alignment file in the bam format (Li et al., 2009), swith duplicated reads marked with Picard tools (v.1.11) was generated for the downstream analyses. We used custom JAVA programs to generate coverage profiles from the alignment bam file. We accepted reads with a BWA read quality of at least 5 and allowed reads that represented duplicated DNA fragments. Resulting files in 'bigWig' format (Kent et al., 2010) were used for visualization in the UCSC browser.

RPKM, data normalization and Z-score

To quantify expression levels using generated genome wide read-profiles we calculated Reads Per Kilobase per Million mapped reads (RPKM) (Mortazavi et al., 2008; Pepke et al., 2009) for all exons annotated in Ensembl v61 (253,750 in total) using the following expression:

$$RPKM = n/(L*N)$$

Where n is a fractional number of reads aligned to the region, L is the length of the region in Kb, and N is the number reads aligned into all annotated exons, in millions. We excluded reads aligned to the mitochondrial genome in calculating the normalization constant, N. For pair-wise sample comparisons, an empirical Z-score was calculated assuming the mean coverage of exon followed a Poisson distribution:

$$Z = (C_A - C_B x_{AB}) / \sqrt{C_A + C_B x_{AB}^2} ,$$

where C_A and C_B are mean coverages of a given exon for samples A and B, respectively, and x_{AB} is a ratio of total number of aligned exonic reads for two samples, $x_{AB} = N_A/N_B$. We also calculated an asymmetry as

$$A = (C_A - C_B x_{AB}) / (C_A + C_B x_{AB})$$

and a fold-change

$$FC = (1 + A) / (1 - A)$$

We first calculated the Z-score, asymmetry and fold-change for every individual exon. Exons that had $Z \geq 2$ and $FC \geq 1.5$ were considered as upregulated, while exons that had $Z \leq 2$ and $FC \leq 0.66$ were considered downregulated between samples.

Alternatively we also calculated all above observables for the whole gene. In this case, the mean coverage was calculated as the sum of total coverage for all exons in

the gene divided by total exonic length. The same thresholds for the z-score and fold-change were applied to determine lists of up- and downregulated genes.

CHIP-Seq analysis

Single-end 100 bp sequence reads were mapped to hg19 (NCBI 37) using Burrows-Wheeler aligner (BWA) (v0.5.9) (Li et al., 2009) and default parameters. Reads having identical coordinates were collapsed into a single read and reads with BWA alignment quality ≥ 5 (in this way, we also filtered reads aligned into multiple locations) were directionally extended 150 basepairs. By applying custom JAVA software to extended ChIP-seq reads we generated genome-wide profiles in the Wig and bigWig formats (Kent et al,. 2010) to be visualized in the UCSC genome browser (Karolchik et al., 2007) and for downstream analysis.

Different approaches were applied to the analysis of various epigenetic modifications. To determine genes that were marked with H3K36me3, we calculated read coverage and RPKM over the annotated gene models (Ensembl v61). By considering genome-wide distribution of H3K36me3, and comparing RPKM calculated for the random genomic regions, we selected a threshold to separate the signal from the background, and by doing so, determined the genes that are marked by this histone modification. To determine genes that were marked with a H3K4me3 mark, we calculated RPKM for the regions flanking transcription start sites (TSS) (+/- 1000 bp) and used FindPeaks2 software (Fejes et al., 2008) to identify enriched locations at FDR ~0.01. To determine genes that were marked with a H3K27me3 mark, we calculated RPKM for the regions flanking transcription start sites (TSS) (+/- 1000 bp) and used the genome-wide H3K27me3 distribution to determine a threshold discriminating marked promoters from unmarked promoters.

RESULTS

Histone modifications in proximal tubular epithelial cells

Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) can be used genome-wide to identify genomic segments that are bound by chromatin-binding proteins or carry a specific histone modification. Here, we aimed to identify genes that have a H3K36me3 mark in proximal tubular epithelial cells. We validated the ChIP-seq method on chromatin from HEK293T cells. Chromatin immuno-precipitation using an antibody specific for H3K36me3 resulted in an approximately 6.0-fold enrichment of genomic segments that have a H3K36me3 mark, as compared to immunoprecipitation using an antibody against the core of histone 3 (H3). Subsequent next-generation sequencing of the DNA-content of the precipitated chromatin revealed H3K36me3 profiles that showed similarity to previously published H3K36me3 profiles (Barski et al., 2007; Mikkelsen et al., 2007) (*data not shown*). Having validated the method, we then performed ChIP-seq experiments on two proximal tubular epithelial cell samples to identify the set of genes that have a H3K36me3 mark in this cell type. In parallel, we assessed the H3K4me3 and H3K27me3 profiles

in one of the proximal tubular epithelial cell samples. The enrichment for the ChIP-seq experiments on chromatin from the proximal tubular epithelial cells was 4.8, 7.0, and 12.3 for H3K36me3, H3K27me3, and H3K4me3, respectively. Subsequent next-generation sequencing (for ChIP-seq alignment statistics, see *Supplementary Table S1*) revealed a distribution of the histone modifications across the genome that was expected from literature: H3K36me3 is mainly present on the "body" of active genes. H3K27me3, which is associated with inactive genes, shows a reciprocal pattern with H3K36me3. H3K4me3 is enriched around transcription start sites of active or poised genes (*Figure 1*). The differences in distribution of the histone modifications explain the differences in enrichment following chromatin immunoprecipitation; in contrast with H3K36me3 and H3K27me3, which are located at broad genomic regions, H3K4me3 is located at much more restricted regions, and therefore shows a higher enrichment. We compared the distribution of H3K36me3, H3K27me3, and H3K4me3 in proximal tubular epithelial cells with previously published ChIP-seq data from adult kidney and fetal kidney samples (GEO database GSM621634 and GSM621634, respectively). Overall, the patterns show a striking resemblance. However, for several genes that are known to be specifically expressed in endothelial cells we see H3K4me3 and H3K36me3 marks in the "total kidney" samples, which contain endothelial cells, while these marks are absent in proximal tubular epithelial cells (*Figure 2*), showing that a subset of genes has tissue-specific histone marks.

We correlated the level of histone modifications present on genes with the transcript levels of these genes in the proximal tubular epithelial sample, as assessed by RNA-seq (For RNAseq alignment statistics, see *Supplementary Table 2*). For H3K4me3 and H3K36me3, we clearly see a positive correlation with transcript levels. For H3K27me3, we clearly see a negative correlation (*Figure 3*).

siRNA-mediated knockdown of SETD2

SETD2 transcript levels were decreased approximately 60% in siRNA-treated PTECs compared to wildtype PTECs (*Figure* 4.a). This resulted in a decrease of the level of global H3K36me3 of approximately 70%, as depicted in *Figure 4b*. We included clear cell renal cell carcinoma cell line RCC-AB in this experiment. RCC-AB harbors an inactivating mutation in *SETD2*, which results in loss of H3K36me3 (Duns et al., 2010). The observed faint residual staining is probably an effect of nonspecific binding of the antibody (*Figure 4b*). The effect of the decreased H3K36me3 levels in proximal tubular epithelial cells on expression and alternative splice site choice of target genes was assessed by RNA-seq.

Effect of *SETD2* knockdown on transcript levels

We observed a significant decrease (fold-change (FC) <2/3, z-score<-2.0) in expression for 55 protein coding genes, and a significant increase in expression for 33 protein coding genes as a result of *SETD2* knockdown (FC>1.5, z-score>2.0) (*Table 1*). As expected, *SETD2* is in the list of genes that show a significant decrease in transcript level. The differentially regulated genes were analyzed for functional enrichment using *DAVID* (Dennis et al., 2003). The Functional Annotation Chart Tool was used to obtain

Table 1. Genes showing significantly altered transcript levels as a result of *SETD2* knockdown in proximal tubular epithelial cell sample PTEC-2A09.

Name	Ensembl_ID	RPKM (WT)	RPKM (SETD2 KD)	Fold change
EIF4EBP1	ENSG00000187840	2649	1020	-2.6
BNIP3	ENSG00000176171	9552	3886	-2.5
ANGPTL4	ENSG00000167772	5124	2166	-2.4
ENO2	ENSG00000111674	7480	3235	-2.3
ADAMTS5	ENSG00000154736	1435	621	-2.3
TFPI2	ENSG00000105825	14229	6379	-2.2
PLAT	ENSG00000104368	8512	3881	-2.2
IGFBP6	ENSG00000167779	1907	903	-2.1
INSIG2	ENSG00000125629	1920	910	-2.1
PRDX3	ENSG00000165672	9333	4578	-2.0
NHP2L1	ENSG00000100138	3003	1483	-2.0
PGK1	ENSG00000102144	36735	18305	-2.0
SETD2	ENSG00000181555	2172	1084	-2.0
TMCO1	ENSG00000143183	2914	1460	-2.0
DDIT4	ENSG00000168209	3589	1817	-2.0
AKAP12	ENSG00000131016	26489	13873	-1.9
WBP2	ENSG00000132471	3956	2072	-1.9
CRKL	ENSG00000099942	2490	1311	-1.9
HK2	ENSG00000159399	3392	1788	-1.9
C4orf3	ENSG00000164096	4410	2341	-1.9
NDRG1	ENSG00000104419	3933	2128	-1.8
P4HA1	ENSG00000122884	4546	2575	-1.8
FAM162A	ENSG00000114023	4358	2483	-1.8
PLAU	ENSG00000122861	41615	23798	-1.7
C19orf33	ENSG00000167644	7727	4459	-1.7
LDHA	ENSG00000134333	66617	38508	-1.7
ASNS	ENSG00000070669	3413	1982	-1.7
RPS23	ENSG00000186468	10212	5937	-1.7
IMPAD1	ENSG00000104331	3471	2048	-1.7
CCND1	ENSG00000110092	32953	19560	-1.7
TPI1	ENSG00000111669	11127	6644	-1.7
ADAMTS1	ENSG00000154734	16711	10129	-1.6
MYADM	ENSG00000179820	3876	2367	-1.6
MARCKS	ENSG00000155130	3198	1959	-1.6
C12orf57	ENSG00000111678	3361	2067	-1.6
ERO1L	ENSG00000197930	5763	3557	-1.6
BNIP3L	ENSG00000104765	7439	4597	-1.6
SERPINE1	ENSG00000106366	39433	24618	-1.6
LGALS1	ENSG00000100097	15285	9543	-1.6
CTGF	ENSG00000118523	18976	11958	-1.6
SLC2A1	ENSG00000117394	10726	6863	-1.6
C10orf55	ENSG00000222047	5515	3546	-1.6
KRT8	ENSG00000170421	15711	10138	-1.5
PRNP	ENSG00000171867	4746	3064	-1.5
LAMC2	ENSG00000058085	42671	27688	-1.5
THBS1	ENSG00000137801	22046	14346	-1.5

Table 1 continued.

Name	Ensembl_ID	RPKM (WT)	RPKM (SETD2 KD)	Fold change
PSAT1	ENSG00000135069	6079	3970	-1.5
RPS29	ENSG00000213741	9136	5975	-1.5
TNFRSF12A	ENSG00000006327	7716	5053	-1.5
HMGA1	ENSG00000137309	19310	12666	-1.5
TGFA	ENSG00000163235	7120	4724	-1.5
GPX8	ENSG00000164294	3722	2472	-1.5
RNF187	ENSG00000168159	4856	3264	-1.5
GTF2H5	ENSG00000185068	4528	3046	-1.5
KIF1C	ENSG00000129250	4529	3086	-1.5
DEFB1	ENSG00000164825	55	542	9.9
C2CD4A	ENSG00000198535	97	757	7.8
SAA2	ENSG00000134339	882	3295	3.7
GPX3	ENSG00000211445	871	2658	3.1
VGF	ENSG00000128564	645	1876	2.9
C3	ENSG00000125730	842	2402	2.9
SAA1	ENSG00000173432	3129	8099	2.6
C1R	ENSG00000159403	809	1937	2.4
SST	ENSG00000157005	742	1704	2.3
CDC25B	ENSG00000101224	965	2138	2.2
IFI6	ENSG00000126709	1727	3813	2.2
DEPDC6	ENSG00000155792	981	2133	2.2
GBP2	ENSG00000162645	973	2087	2.1
MORC4	ENSG00000133131	3394	6539	1.9
MAOA	ENSG00000189221	2142	4083	1.9
MMP7	ENSG00000137673	2738	5045	1.8
ABCA1	ENSG00000165029	1921	3516	1.8
CEBPD	ENSG00000221869	1708	3069	1.8
WDR74	ENSG00000133316	56790	101306	1.8
CPD	ENSG00000108582	10429	18217	1.7
NRBP2	ENSG00000185189	4067	7039	1.7
CRYM	ENSG00000103316	1965	3332	1.7
ABCC3	ENSG00000108846	2822	4778	1.7
TMEM2	ENSG00000135048	2798	4692	1.7
SOD2	ENSG00000112096	10925	17992	1.6
PROM1	ENSG00000007062	2874	4635	1.6
NAMPT	ENSG00000105835	2761	4409	1.6
SLC34A2	ENSG00000157765	4084	6446	1.6
SYNE2	ENSG00000054654	4104	6270	1.5
SAT1	ENSG00000130066	7546	11412	1.5
TNFSF10	ENSG00000121858	5743	8669	1.5
GLS	ENSG00000115419	10013	15094	1.5
TNFRSF11B	ENSG00000164761	7953	11676	1.5

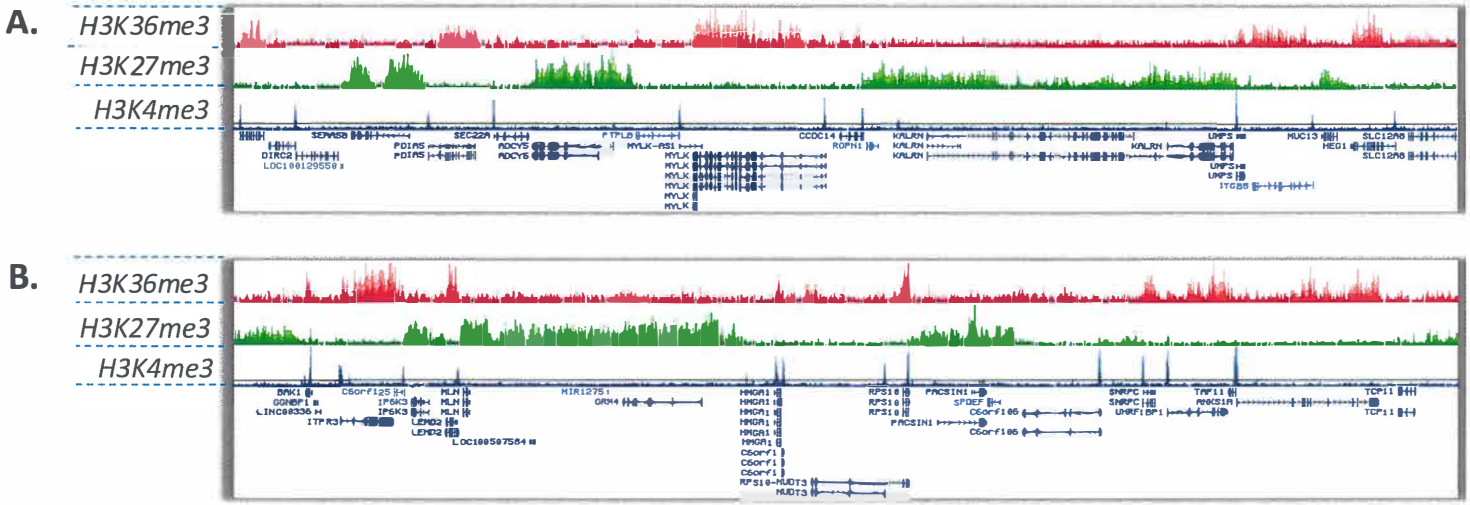


Figure 1. Histone modification profiles in proximal tubular cell sample PTEC-2A09. Coverage profiles for H3K36me3 (top), H3K37me3 (middle) and H3K4me3 (bottom) generated from the bam alignment file (see Materials and Methods), and displayed as custom tracks on the UCSC genome browser. The profiles represent the number of sequence reads uniquely aligning to (**A**) segments of chromosome 3 (122.38-124.90 Mb from pter), and (**B**) chromosome 6 (33.43-35.17 Mb from pter). The bottom part of the figures shows the locations of the RefSeq genes, with boxed regions and thin lines representing exons and introns, respectively.

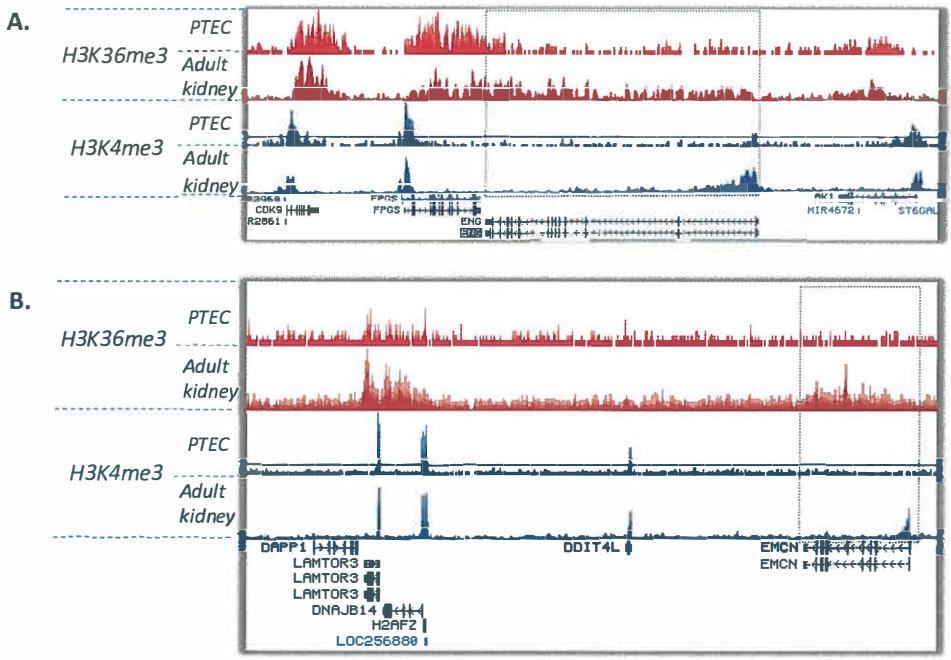


Figure 2. H3K36me3 and H3K4me3 profiles on endothelium-specific genes. Coverage profiles for H3K36me3 and H3K4me3 in proximal tubular epithelial cell sample PTEC-2A09 (PTEC) and an adult kidney sample (GEO database GSM621634), generated from the bam alignment file (see Materials and Methods), and displayed as custom tracks on the UCSC genome browser. The profiles represent the number of sequence reads uniquely aligning to segments of (**A**) chromosome 9 (130.55-130.66 Mb from pter), and (**B**) chromosome 4 (100.6-101.6 Mb from pter). The bottom parts of the figures show the location of the RefSeq genes in these genomic segments, with boxed regions and thin lines representing exons and introns, respectively. The *ENG* and *EMCN* genes, which code for the endothelial markers Endoglin and Endomucin, respectively, are highlighted by dotted squares in **A** and **B**. These genes show H3K36me3 and H3K4me3 marks in the total kidney sample, which are largely absent in the proximal tubular epithelial cells (PTECs).

the p-values that describe the probability that a functional term will be enriched among the up- or downregulated genes. Gene set enrichment analysis for KEGG pathways for the set of downregulated genes reveals enrichment for, amongst others, the glucolysis/ gluconeogenesis pathway, the p53 signaling pathway, renal cell carcinoma, and the ErbB signaling pathway, amongst others (*Table 2*). Gene set enrichment analysis for GO terms for the set of downregulated genes reveals enrichment for angiogenesis related processes, amongst others (*Table* 2). Gene set enrichment analysis for KEGG pathways for the set of upregulated genes reveals enrichment for arginine and proline metabolism. Gene set enrichment analysis for GO terms for the set of upregulated genes reveals enrichment for homeostatic processes (*Table 2*).

Table 2. Enriched KEGG pathways and top 10 enriched GO terms for genes showing significantly altered transcript levels as a result of *SETD2* knockdown.

	Downregulated genes					Upregulated genes				
	Term	# genes	%	P-value	Benjamini-Hochberg	Term	# genes	%	P-value	Benjamini-Hochberg
Enriched KEGG pathways	Glycolysis/Gluconeogenesis	5	8.5	3.80E-04	2.20E-02	Arginine and proline metabolism	3	8.8	7.70E-03	1.80E-01
	P53 signaling pathway	3	5.1	6.00E-02	8.40E-01					
	Complement and coagulation cascades	3	5.1	6.20E-02	7.10E-01					
	Renal cell carcinoma	3	5.1	6.30E-02	6.10E-01					
	ErbB signaling pathway	3	5.1	9.20E-02	6.80E-01					
	Ribosome	3	5.1	9.20E-02	6.80E-01					
	Prostate cancer	3	5.1	9.60E-02	6.20E-01					
GO terms (BP_FAT)	Blood vessel development	8	13.6	2.00E-05	1.50E-02	chemical homeostasis	7	20.6	3.40E-04	2.00E-01
	Glycolysis	5	8.5	2.00E-05	7.90E-03	homeostatic process	7	20.6	2.50E-03	5.50E-01
	Vasculature development	8	13.6	2.30E-05	6.00E-03	cellular ion homeostasis	5	14.7	5.30E-03	6.80E-01
	glucose catabolic process	5	8.5	4.70E-05	9.10E-03	cellular ion homeostasis	5	14.7	5.60E-03	5.90E-01
	blood vessel morphogenesis	7	11.9	8.40E-05	1.30E-02	ion homeostasis	5	14.7	7.30E-03	6.00E-01
	hexose catabolic process	5	8.5	9.40E-05	1.20E-02	response to drug	4	11.8	8.00E-03	5.70E-01
	response to hypoxia	6	10.2	9.70E-05	1.10E-02	regulation of inflammatory response	3	8.8	9.30E-03	5.70E-01
	monosaccharide catabolic process	5	8.5	1.00E-04	1.00E-02	cellular homeostasis	5	14.7	1.10E-02	6.00E-01
	response to oxygen levels	6	10.2	1.20E-04	1.10E-02	behavior	5	14.7	1.20E-02	5.60E-01
	Arginine and proline metabolism	3	8.8	7.70E-03	1.80E-01	lymphocyte chemotaxis	2	5.9	1.30E-02	5.80E-01

Table 3. Genes showing a change in alterative splice site choice as a result of *SETD2* knockdown.

			Wild type				SETD2 knock down			
Gene Symbol	Ensemble ID	Alternatively skipped exon	# reads retained	skipped	fraction retained	skipped	# reads retained	skipped	fraction retained	skipped
AASS	ENSG00000008311	19	4	9	0.3	0.7	4	2	0.7	0.3
DPM1	ENSG00000000419	7	0	11	0.0	1.0	7	17	0.3	0.7
LTBP3	ENSG00000168056	24	5	3	0.6	0.4	34	8	0.8	0.2
PUM2	ENSG00000055917	2	1	9	0.1	0.9	25	37	0.4	0.6
RAD1	ENSG00000113456	3	2	11	0.2	0.8	5	3	0.6	0.4
AXIN1	ENSG00000103126	9	11	3	0.8	0.2	1	1	0.5	0.5
RCC1	ENSG00000180198	6'	9	9	0.5	0.5	2	3	0.4	0.6

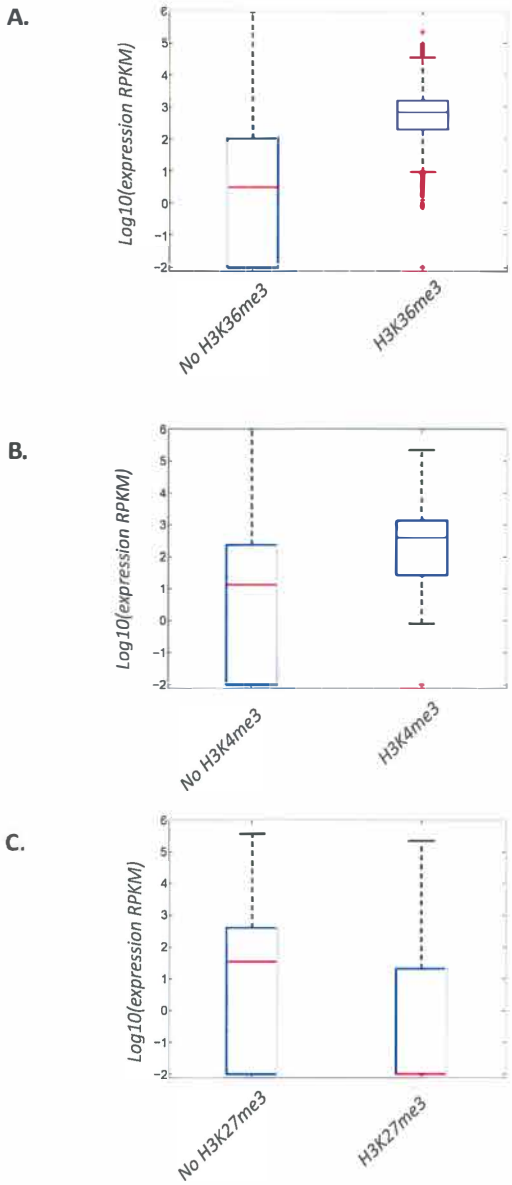


Figure 3. Correlation between the presence of histone modifications and transcript levels in proximal tubular sample PTEC-2A09. Box plots showing the distribution of transcript levels (log_{10}RPKM) for all Ensembl v61 human transcripts in untreated PTEC-2A09 cells, as assessed by RNA-seq. (**A**) **H3K36me3.** The H3K36me3 coverage was measured as the total number of reads falling within gene boundaries normalized by the length of the gene. Using the cumulative distributive function (CDF), an empirical threshold was selected to split genes into subsets having H3K36me3 or no H3K36me3. (**B**) **H3K4me3**. The H3K4me3 coverage was measured as the total number of reads falling within 1 Kb up- and downstream of the transcription start site (TSS), normalized by 2 Kb. Using the cumulative distributive function (CDF), an empirical threshold was selected to split genes into subsets having H3K4me3 or no H3K4me3. (**C**) **H3K27me3**. The H3K27me3 coverage was measured as the total number of reads falling within 1 Kb up and downstream of the transcription start site (TSS), normalized by 2 Kb. Using the cumulative distributive function (CDF), an empirical threshold was selected to split genes into subsets having H3K27me3 or no H3K27me3.

Effect of *SETD2* knockdown on alternative splice site choice

Numerous individual exons show a significant decrease (FC ≤ -1.5, z-score ≤ -2.0) or increase (FC ≥ 1.5, z-score ≥ 2.0) of transcript levels as a result of SETD2 knockdown, whereas the transcript level for the whole gene is not significantly changed. This might be the result of a change in alternative splice site choice. We performed a preliminary data analysis, and focused on genes that showed significantly altered levels for one

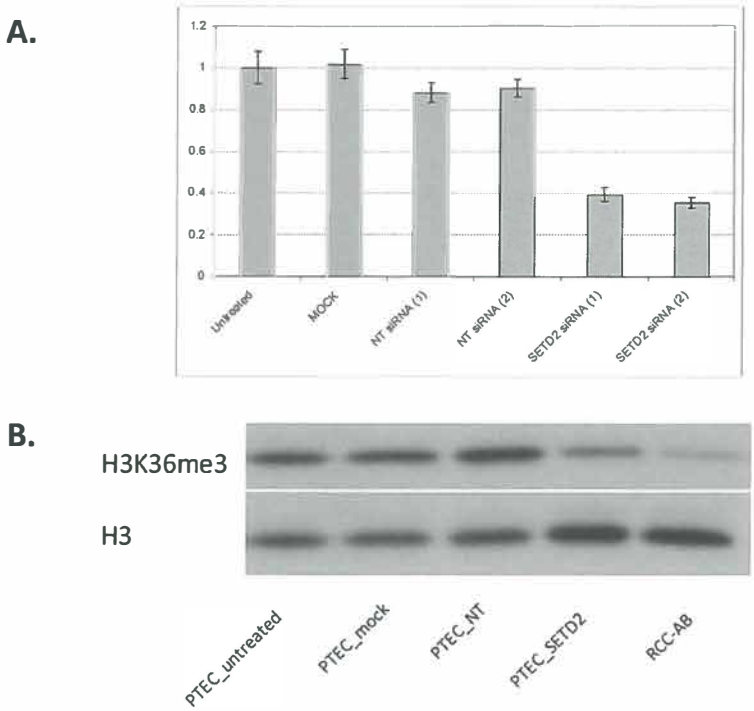
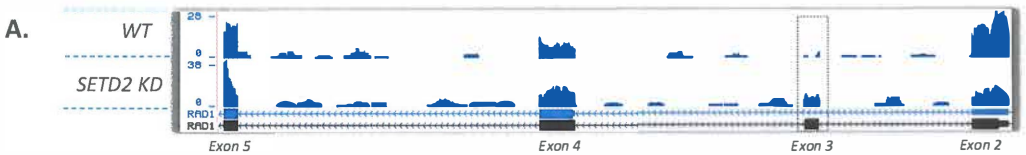
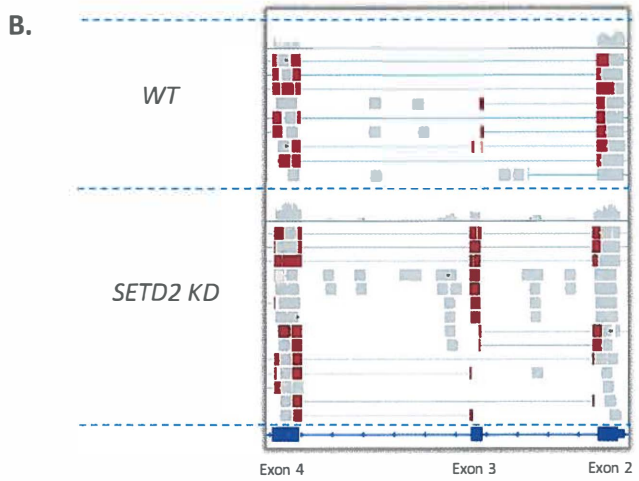


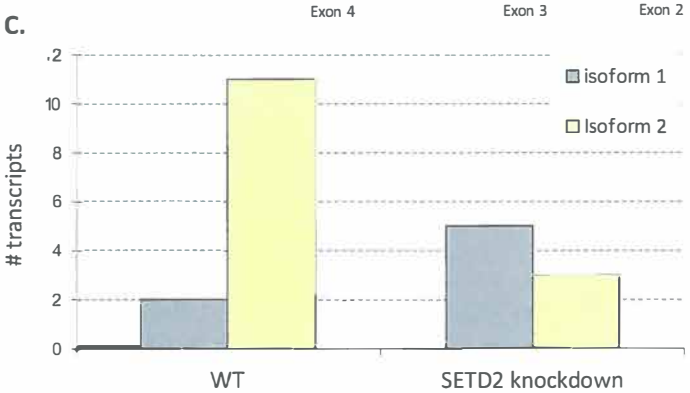
Figure 4. siRNA-mediated knockdown of SETD2 in proximal tubular epithelial cells. (A) SETD2 transcripts levels relative to the level in untreated cells, as assessed by RT-PCR. (B) H3K36me3 levels as assessed by Western blot analysis. Blotting with anti-histone 3 (H3) served as a loading control.

exon. Individual reads mapping at, and around, the alternatively spliced exons were visualized using the *Integrative Genomics Viewer* (*IGV*) (http://www.broadinstitute.org/igv/). Subsequently, the ratios between the number of reads that skip the exon of interest and the number of reads that retain the exon of interest were assessed, and compared between the wildtype (WT) sample and the SETD2 knockdown (SETD2 KD) sample. Our preliminary data analysis revealed possible changes in alternative splice site choice as a result of *SETD2* knockdown for *AASS*, *Axin 1*, *DPM1*, *LTBP3*, *PUM2*, *RAD1*, and *RCC1* (*Table 3*). In *Figure 5*, the transcript levels for the different isoforms of *RAD1* are shown. *SETD2* knockdown appears to result in a shift from isoform 2 (which lacks exon 3) towards isoform 1 (which retains exon 3) of *RAD1*. All the genes that show differential exon usage as a result of *SETD2* knockdown have a H3K36me3 mark in the wildtype PTEC sample, as is shown for *RAD1* in *Figure 5d*.

Besides the apparent change in splice site choice for some genes, we observed an apparent change in transcriptional start site usage as a result of decreased H3K36me3 levels for several genes, including *STK19*. Knockdown of *SETD2* appears to result in usage of a downstream alternative start site of *STK19* (*Figure 6*).







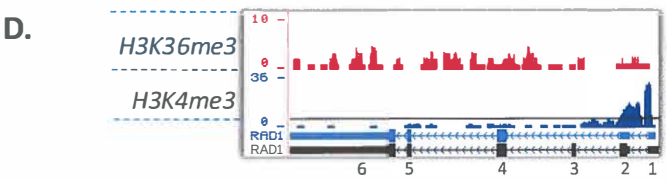


Figure 5. Differential alternative splice site choice as a result of *SETD2* knockdown in proximal tubular epithelial cell sample PTEC-A09. (**A**) Coverage profiles of transcript levels, generated from the bam alignment file (see Materials and Methods), and displayed as custom tracks on the UCSC genome browser. The profiles represent the number of sequence reads uniquely aligning to the *RAD1* gene, as assessed by RNA-seq. The exon-intron structure of *RAD1* (RefSeq) is shown at the bottom, with boxed regions and thin lines representing exons and introns, respectively. *WT*: untreated PTEC sample, SETD2 KD: PTEC sample treated with *SETD2* siRNA. (**B**) Individual sequence reads mapping to *RAD1*, as visualized by the *Integrative Genomics Viewer*. Boxes represent the aligned sequence reads. Total coverage is shown on top. Reads that span different exons are represented by two or more separate boxes connected by thin lines. In this example, several sequence reads skip exon 3 of *RAD1* in the WT sample, while most of the sequence reads in the SETD2 knockdown sample show retainment of this exon. (**C**) Graph representing the number of sequence reads that show exclusion (isoform 2) or retainment (isoform 1) of exon 3 of *RAD1*. (**D**) Coverage profiles for H3K36me3 (top) and H3K4me3 (bottom) generated from the bam alignment file (see Materials and Methods), and displayed as custom tracks on the UCSC genome browser. The profiles represent the number of sequence reads uniquely aligning to the *RAD1* gene. The exon-intron structure of *RAD1* (RefSeq) is shown at the bottom, with boxed regions and thin lines representing exons (numbered) and introns, respectively.

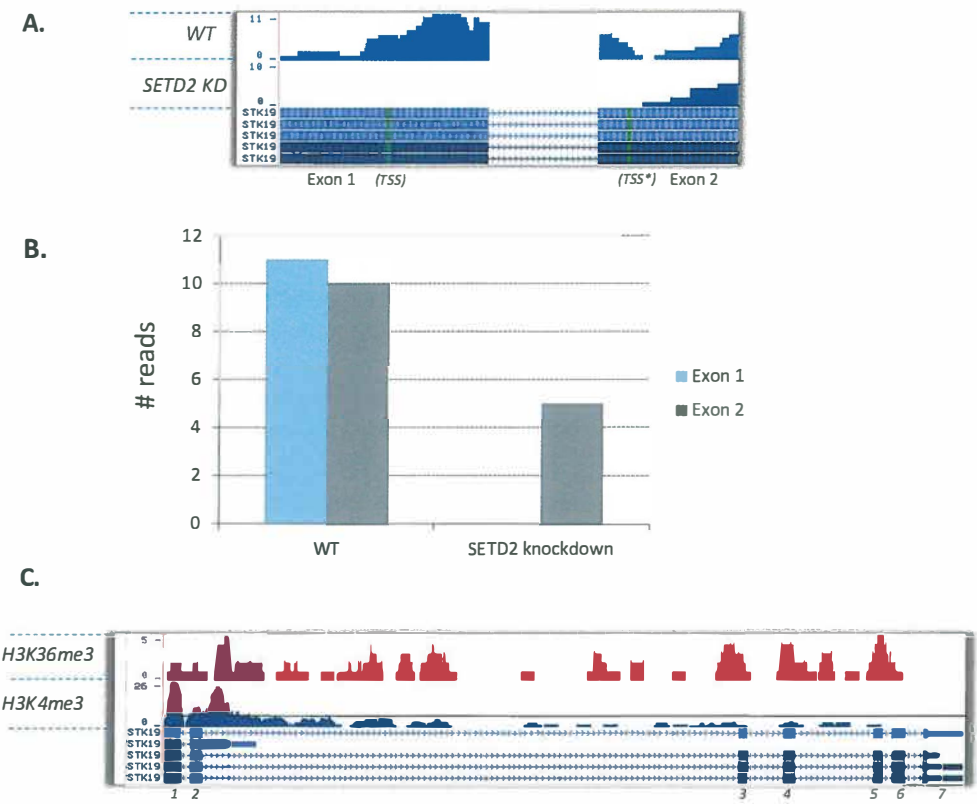


Figure 6. Differential alternative start site usage *STK19* as a result of *SETD2* knockdown in proximal tubular epithelial cell sample PTEC-2A09. (**A**) Coverage profiles of transcript levels, generated from the bam alignment file (see Materials and Methods), and displayed as custom tracks on the UCSC genome browser. The profiles represent the number of sequence reads uniquely aligning to the first part of the *STK19* gene, as assessed by RNA-seq. The exon-intron structure of the first part of *STK19* (RefSeq) is shown at the bottom, with boxed regions and thin lines representing exons (numbered) and introns, respectively. TSS: transcription start site, TSS*: alternative transcription start site. *WT*: untreated PTEC sample, SETD2 KD: PTEC sample treated with *SETD2* siRNA. (**B**) Graph representing the number of sequence reads that map to exon 1 or exon 2 of *STK19* in the untreated and *SETD2* knockdown samples. (**C**) Coverage profiles for H3K36me3 (top) and H3K4me3 (bottom) generated from the bam alignment file (see Materials and Methods), and displayed as custom tracks on the UCSC genome browser. The profiles represent the number of sequence reads uniquely aligning to the *STK19* gene. The exon-intron structure of *STK19* (RefSeq) is shown at the bottom, with boxed regions and thin lines representing exons (numbered) and introns, respectively.

H3K36me3 in different tissue types

We used several publically available data sets from ChIP-seq experiments (including GEO databases GSM537699 (adult liver), GSM613873 (breast myoepithelial cells, and GSM669597 (left luminal epithelium) to assess whether H3K36me3 modification of the investigated genes was specific for proximal tubular epithelium. None of the above genes appeared to have a H3K36me3 mark specifically in proximal tubular epithelial cells (*data not shown*).

DISCUSSION

Inactivation of *SETD2* has been suggested to be specific for clear cell renal cell carcinoma (Dalgliesh et al., 2010). This might indicate the presence of H3K36me3 on a subset of genes specifically in proximal tubular epithelial cells. A change in transcription or transcription-related processes for these genes as a result of H3K36me3 loss might contribute to ccRCC development. As expected, the H3K36me3 profiles in PTECs are largely similar to those in other tissues. However, adult kidney and proximal tubular epithelial cells did show differences in their H3K36me3 profiles for a small subset of genes. Some of these genes are specifically expressed in endothelium cells, that are present in adult kidney, showing that a subset of genes has tissue-specific histone marks. Comparison of the H3K36me3 profiles of PTECs with other epithelial cells may identify a set of genes that specifically carry a H3K36me3 mark in PTECs. As an initial step, the analysis of bronchial epithelial cells will be carried out.

Transient siRNA-mediated knockdown of *SETD2* in proximal tubular cells, which resulted in a significant decrease in global H3K36me3 levels (*Figure 4*), results in significantly increased transcript levels of 33 genes and significantly decreased transcript levels of 55 genes. The set of downregulated genes showed a significant enrichment for genes that function in angiogenesis-related processes (*Table 2*). This is in concordance with the observation that SETD2 is required for embryonic vascular remodeling (Hu et al., 2009). The enrichment of genes that are involved in the p53 signaling pathway might imply that *SETD2* does indeed function in regulating a subset of P53 target genes (Xie et al., 2009). For most of the genes, it remains unclear if the observed changes in transcript levels are a direct result of *SETD2* knockdown. The observation that H3K36me3 is virtually absent on *DEFB1* and *C2CD4A* in the PTEC sample implies that the observed increased transcript levels for these genes are not the result of loss of H3K36me3.

Several individual exons show a significant decrease of transcript levels as a result of SETD2 knockdown, whereas the transcript level for the whole gene is not significantly changed. This might be the result of a change of an alternative splice site choice. Our preliminary data analysis revealed an apparent change in alternative splice site choice as a result of *SETD2* knockdown for seven genes (*Table 3*). This indeed suggests the involvement of H3K36me3 in alternative splice site choice for at least a subset of genes. However, we did not confirm the changes in alternative splicing as a result of the modulation of H3K36me3 levels for *FGFR2*, *TPM1*, *TPM2*, and *PKM2* that were observed by Luco et al. (2010) in this study. *TPM1* and *PKM2* are in our list of genes that show significant changes in the expression of an individual exon, while not showing differential transcript levels for the whole gene. This might indicate that alternative splice site choice for these genes has indeed changed as a result of *SETD2* knockdown. However, due to numerous reads mapping to the intronic regions of these genes, we could not reliably assess the numbers of reads skipping or retaining the exons concerned. *FGFR2* did not show detectable transcript levels, and *TPM2* did not show any detectable change in alternative exon usage *(data not shown)*.

Our preliminary data analysis also revealed a set of genes for which H3K36me3 downregulation appears to result in change of their transcriptional start site. The observed changes should be treated with caution, as most of the transcripts show a lower coverage towards the 5' end, probably as a result of mRNA degradation, which in most cases starts in the 5' region of an mRNA molecule. The apparent changes in transcriptional start site usage could therefore be false-positives, due to low coverage. Either way, a possible link between H3K36me3 and intragenic transcription in mammals was recently provided by Xie et al. (2011). They showed that recruitment of histone demethylase KDM5B to H3K36me3 via interaction with chromodomain protein MRG15 decreased intragenic H3K4me3, resulting in suppression of cryptic intragenic transcription (Xie et al., 2011).

Several of the genes that show changes in alternative exon usage are interesting with respect to tumorigenic development. LTBP3 plays a critical role in TGF-β signaling (Hyytiainen et al., 2004) Isoform 1 and 2 of *LTBP3* are predominantly expressed in different tissue types, suggesting different functions in cellular regulation for these different isoforms. Axin 1 is a negative regulator of WNT-signaling. The observed shift from isoform 1 towards isoform 2 of Axin 1 has been previously observed in ovarian tumors (Klinck et al., 2008). *PUM2* and *RAD1* code for proteins that control cell cycle entry (Kedde et al., 2010) and play a role in DNA repair (Zhang et al., 2011), respectively. So, given their function, deregulation of several of these genes could potentially contribute to ccRCC development. At the moment, knockdown experiments in additional PTEC samples are ongoing to confirm the role of SETD2/H3K36me3 in the choice of alternative splice site or start site for these genes. Additional studies will be needed to assess the relevance of these altered splicing events with respect to ccRCC development. Comparing the abundance of different isoforms between large numbers of ccRCC tumor samples and normal renal epithelial samples could provide clues about the fraction of ccRCC tumors in which these altered splicing events play a role. The genes for which we observed a change of alternative splice site selection or a change in alternative transcription start site all have a H3K36me3 mark in proximal tubular epithelial cells. However, the mark appeared not to be specific for proximal tubular epithelial cells, as H3K36me3 is also present on these genes in several other tissue types, including breast epithelium. There might, however, be more subtle differences in the local amount of H3K36me3 levels between proximal tubular epithelial cells and the other tissue types. The resolution of our H3K36me3 ChIP-seq data is not sufficient to detect these subtle differences. Another possibility is that, although the gene carries a H3K36me3 mark in several tissues, the change in exon usage has a tumorigenic effect in proximal tubular epithelial cells, but not in other cell types. Either way, additional data are being collected to further analyze H3K36me3 positions in PTEC samples and to compare them with those in other tissue types, such as bronchus epithelial cells.

H3K36me3 is present on the "body" of genes that are transcribed by RNA Polymerase II, i.e., a majority of actively transcribed protein coding genes (Edmunds et al., 2008). However, *SETD2* knockdown in proximal tubular cells appears to affect the processing of only a limited number of target genes. One reason might be that the RNAi-mediated knockdown of SETD2 was not sufficient. Although the global

H3K36me3 level appears to be strongly decreased (*Figure 4*), the residual H3K36me3 levels might be sufficient to result in the "normal" processing of most of the target genes. The ongoing experiments should reveal whether a stronger and/or prolonged knockdown of SETD2 and H3K36me3 levels indeed has a stronger effect on transcript levels, alternative splice site choice, or choice of transcriptional start site for numerous additional target genes.

ACKNOWLEDGEMENTS

We thank Jackie Senior for editing the manuscript.
This project has been financially supported by the Dutch Cancer Society (Grant 2007-3892) and the Foundation "De Drie Lichten" in the Netherlands.

REFERENCES

Allemand E, Batsché E, Muchardt C. Splicing, transcription, and chromatin: a ménage à trois. Current opinion in Genetics and development 2008:18;145.

Al Sarakbi W, Sasi W, Jiang WG, et al. The mRNA expression of SETD2 in human breast cancer: correlation with clinico-pathological parameters. BMC Cancer 2009;9;290.

Andersson R, Enroth S, Rada-Iglesias A, et al. Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Research 2009;19:1732-41.

Barski A, Cuddapah S, Cui K, et al. High- resolution profiling of histone methylations in the human genome. Cell 2007;129:823.

Cáceres JF, Kornblihtt AR. Alternative splicing: multiple control mechanisms and involvement in human disease. Trends in Genetics 2002;18:186.

Cooper TA, Wan L, Dreyfuss G. RNA and disease. Cell 2009;136:777.

Dalgliesh GL, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying enzymes. Nature 2010:463;360.

De la Mata M, Alonso CR, Kadener S, et al. A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell 2003;12:525.

Dennis G Jr, Sherma BT, Hosacj DA, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003;4:P3.

Duns G, Van den Berg E, Van Duivenbode I, et al. Histone Methyltransferase Gene SETD2 Is a Novel Tumor Suppressor Gene in Clear Cell Renal Cell Carcinoma. Cancer Res 2010;70:4287.

Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/ SETD2 mediates all H3K36 trimethylation. EMBO J 2008;27:406.

Fejes AP, Robertson G, Bilenky M, et al. FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. Bioinformatics 2008;24:1729.

Flicek P., Amode MR, Barrell D, et al. Ensembl 2011. Nucleic Acids Res 2010;39:D800.

Kent WJ, Zweig AS, Barber G, et al. BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 2010;26:2204.

Füllgrabe J, Kavanagh E, Joseph B. Histone onco-modifications. Oncogene 2011;30:3391.

Hu M, Sun X-J, Zhang Y-L, et al. Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. Proc Natl Acad Sci U S A 2010;107:2956.

Huff JT, Plocik AM, Guthre C, et al. Reciprocal intronic and exonic histone modification regions in humans. Nature Structural & Molecular Biology 2010;17:1495.

Hyytiainen M, Penttinen C, Keski-Oja. Latent TGF-beta binding proteins: extracellular matrix association and roles in TGF-beta activation. Crit Rev Clin Lab Sci 2004;41:233.

Karolchik D, Hinrichs AS, Kent WJ. The UCSC Genome Browser. Curr Protoc Bioinf 2007 Chapter 1:Unit 1.4.

Kedde M, Van Kouwenhove M, Zwart W, et al. A Pumilio-induced RNA structure switch

in p27-3'UTR controls miR-221 and miR-222 accessibility. Nature Cell Biology 2010;12:1014.

Klinck R, Bramard A, Inkel L, et al. Multiple Alternative Splicing Markers for Ovarian Cancer. Cancer Res 2008;68:657.

Kolasinska-Zwierz P, Down T, Latorre I, et al. Differential chromatin marking of introns and expressed exons by H3K36me3. Nature Genetics 2009;41:376.

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25: 1754.

Luco RF, Pan Q, Tominaga K, et al. Regulation of alternative splicing by histone modifications. Science 2010;327:996.

Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621.

Morin, R., Bainbridge, M., Fejes, A., et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. BioTechniques 2008;45:81.

Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011;476:298.

Pepke S, Wold B, Mortazavi A. Computation for ChiP-seq and RNA-seq studies. Nat Methods 2009;6:S22.

Piekielko-Witkowska, Wiszomirska H, Wojcicka A, et al. Disturbed expression of splicing factors in renal cancer affects alternative splicing of apoptosis regulators, oncogenes, and tumor suppressors. PloS One 2010;5:e13690.

Pruitt KD, Maglott DR. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res 2001;29:137.

Rhead B, Karolchik D, Kuhn RM, et al. The UCSC Genome Browser database: update 2010. Nucleic Acids Res 2010;38:D613.

Schor IE, Rascovan N, Pelisch F, et al. Neuronal cell depolarisation induces intragenic chromatin modifications affecting NCAM alternative splicing. Proc Natl Acad Sci U S A 2009;106:4325-30.

Spies N, Nielsen CB, Padgett RA, et al. Biased chromatin signatures around polyadenylation sites and exons. Mol Cell 2009;36:245.

Starr TK, Allaei R, Silverstein KA,et al. A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. Science 2009;323:1747.

Venables JP. Aberrant and alternative splicing in cancer. Cancer Res 2004;64:7647.

Venables JP, Klinck R, Koh C, et al. Cancer-associated regulation of alternative splicing. Nat Struct Mol Biol 2009;16:670.

Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. Nature 2008;456:470.

Xie L, Pelz C, Wang W, et al. KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. EMBO J 2011;30:1473.

Xie P, Tian C, An L, et al. Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. Cell Signal 2008;20:1671.

Yoh SM, Lucas JS, Jones KA. The Iws:Spt6: CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev 2008:22;3422.

Zhang C, Liu Y, Hu Z, et al. Targeted deletion of mouse Rad1 leads to deficient cellular DNA damage responses. Protein cell 2011;2:410.

SUPPLEMENTARY TABLES

Supplementary Table S1. ChIPseq statistics

Sample	mark	total reads	aligned	used	Percent aligned	Percent used
PTEC-2A09	H3K27me3	13352518	9481829	8735566	71.0	65.4
PTEC-2A09	H3K36me3	11988678	8805572	7421825	73.4	61.9
PTEC-2A09	H3K4me3	12041542	7904654	7044424	65.6	58.5
HEK293	H3K36me3	46116644	31150816	28206308	67.5	61.2

Used: reads that passed quality treshhold (QC=5), and duplicated reads are collapsed

Supplementary Table S2. RNAseq statistics

Sample	Total number of reads	Mapped to hg19	Number of reads passing QC	Mapped into Ensembl v61 exons	Mapped into Ensembl v61 genes
PTEC-2A09 (WT)	159681542	136292574	44977120	11118028	20976657
PTEC-2A09 (SETD2 KD)	183622416	156192275	56099498	11956418	27192531

TARGETED EXOME SEQUENCING IN CLEAR CELL RENAL CELL CARCINOMA TUMORS SUGGESTS ABERRANT CHROMATIN REGULATION AS A CRUCIAL STEP IN CCRCC DEVELOPMENT

5

Gerben Duns[1], Robert M.W. Hofstra[1], Jantine G. Sietzema[1], Harry Hollema[2], Inge van Duivenbode[1], Angela Kuik[1], Cor Giezen[1], Jan Osinga[1], Jelkje J. Bergsma[1], Harrie Bijnen[1], Pieter van der Vlies[1], Eva van den Berg[1], Klaas Kok[1].

Departments of [1]Genetics and [2]Pathology & Medical Biology, University Medical Center Groningen, University of Groningen, the Netherlands

Under review

ABSTRACT

Clear cell renal cell carcinomas are characterized by loss of 3p. *VHL*, located at 3p25, is a major ccRCC tumor suppressor gene and is inactivated in the majority of cases. However, inactivation of *VHL* alone is not sufficient for ccRCC development, but rather induces senescence. Recently, *SETD2*, located at 3p21, was identified as a new candidate tumor suppressor gene in ccRCC. Its observed mutational frequency in ccRCC tumors suggests however there are still undiscovered tumor suppressor genes on 3p. We therefore screened all genes on 3p for mutations in ten primary ccRCC tumors using targeted capturing followed by high-througput sequencing. We identified inactivating mutations in *VHL* and *PBRM1* in six and seven out of ten cases, respectively, pinpointing *PBRM1* as a candidate ccRCC tumor suppressor gene. In one primary tumor, we identified a truncating mutation in *BAP1*. Sequencing of *PBRM1* in ccRCC-derived cell lines confirmed its frequent inactivation in ccRCC. *PBRM1* and *BAP1*, which are both located at 3p21, code for BAF180, the chromatin targeting subunit of the SWI/SNF complex, and BRCA1 associated protein-1, which is involved in histone deubiquitination, respectively.

Taken together, these data suggest an important role for aberrant chromatin regulation in ccRCC development.

INTRODUCTION

Clear cell renal cell carcinoma (ccRCC) tumors show a different mutational spectrum compared to most other tumor types. In the majority of cases, the tumor suppressor gene *Von Hippel Lindau* (*VHL*) is inactivated, while classical tumor genes like *P53*, *PTEN*, *KRAS*, *BRAF*, *RB* and *EGFR2* make only a small contribution [Dalgliesh et al., 2010]. Although *VHL* is a major tumor suppressor gene in ccRCC, its inactivation alone does not result in ccRCC development, but rather induces senescence [Mandriota et al., 2002; Young et al., 2003]. This suggests that there are other genes involved in the development of ccRCC.

The main focus in the search for these genes has been on segments of the short arm of chromosome 3 (3p) that show recurrent loss in ccRCC tumors, suggesting the presence of tumor suppressors in these regions. These regions include 3p25, in which *VHL* is located, and 3p21, which is the site of, until recently, unidentified tumor suppressor genes [Van den Berg et al., 1997].

Recently, the histone methyltransferase gene *SETD2*, located at 3p21.31, was identified as a new candidate tumor suppressor gene involved in ccRCC development [Dalgliesh et al, 2010; Duns et al., 2010]. The identification of inactivating mutations in additional genes coding for histone modifying enzymes (*JARID1C*, *UTX*, *MLL2*) in a subset of ccRCC tumors points to a role of aberrant histone modulation in at least a fraction of ccRCC samples [Dalgliesh et al., 2010]. The observed mutational frequency of *SETD2* in ccRCC tumors suggests however that this gene may not be the sole tumor suppressor gene in the 3p21 region. Here we describe the screening of all exons, exon-intron boundaries, and untranslated regions from 3p for mutations in a panel of ten primary ccRCC tumors, using targeted capturing followed by a high-troughput sequencing approach.

MATERIAL AND METHODS

Tumor samples and cell lines

Primary ccRCC tumor samples were snap-frozen in liquid nitrogen, sections were cut and the histology was reviewed by one pathologist (H.Hollema). Tissue biopsies containing at least 70% tumor cells were selected for analysis. The characteristics of these ccRCC-derived cell lines were described earlier [Duns et al., 2010].

SNP analysis

We used the Human CytoSNP-12 BeadChip (Illumina, San Diego, CA, USA) to assess 3p copy number loss in a panel of 13 primary clear cell renal cell carcinoma tumors following the manufacturer's instructions. Data were analysed using NEXUS v6.0 (Biodiscovery, El Segundo, CA, USA).

Exome capture design and library preparation

We developed a custom-designed SureSelect kit (Agilent, Santa Clara, CA, USA) containing, at a 2x coverage, baits for 541 3p genes, UTX, JARID1C, TP53 and ISW (approximately 6000 exons) including their 3'UTRs and 5'UTRs, and for all miRNAs from 3p. The set in total covered 2 Mb of sequences.

Genomic library preparation targeted exome capture

Genomic DNA was isolated from primary ccRCC tumor tissue using the Qiamp DNA Mini Kit (Qiagen, Venlo, The Netherlands) and size-fragmented by Adaptive Focused acoustics on a Covaris E120 (Covaris Inc, Woburn, Massachusetts, USA) (duty cycle: 10%, intensity: 5, cycles per burst: 20, time: 180 s) to a median size of 200 bp. Fragmented DNA was purified using Qiaquick PCR purificaton columns. Genomic libraries were prepared using the NEBNext sample preparation kit following the manufacturer's instructions (NEB, Ipswich, UK). Adaptors were ligated to the end-repaired and dA-tailed DNA fragments using an Illumina Index PE adaptor oligo mix (Illumina). Adapter-ligated DNA fragments were subsequently size- selected (300-350 bp) using E-gel® Clonewell system (Invitrogen, Carlsbad, CA, USA). Size selected adapter-ligated libraries were amplified following the Phusion-high-fidelity-MM protocol (NEB). 500 ng of each amplified adapter-ligated library was hybridized to baits using the SureSelect Target Enrichment System for paired-end multiplex sequencing (Illumina). Hybridized DNA fragments were purified using AMPure beads (Agencourt BioSciences Corporation, Beverley, MA, USA). Index-barcode tags were added by post-hybridization amplification (12-14 cycles) following the Phusion-high-fidelity-MM protocol (NEB). Quality and quantity of fragmented DNA samples and libraries were assessed using Experion DNA 12K analysis kit (Bio-Rad, Hercules, CA, USA).

Sequencing

Enriched barcoded libraries were sequenced with 100 bp paired- end reads across one lane of an Illumina HiSEQ 2000 flowcell (Illumina). The sequence reads were analyzed using the CLC BIO Genomic Work Bench Suite 4.5 (CLC BIO, Arhus, Denmark). After trimming low quality sequences, reads were aligned against chromosome 3 (GRCh37. p2 primary reference assembly).

Variant calling

We used DIP and SNP analysis tools in CLC bio to identify putative indels and single nucleotide changes (quality central base = 20, average quality of surrounding bases = 15, maximum number of gaps and mismatches = 2).

The list of putative variations was shortened by discarding those present in: (1) the dbSNP and/or 1000 Genome Project databases, (2) more than one sample, (3) 100 % of reads, or those that were (4) intergenic, i.e. not in an annotated gene. Furthermore, only putative variants present in paired reads were selected, as these reads have a high probability of being matched correctly. The functional impact of mutations was predicted using Alamut (Interactive Biosoftware, Rouen, France), NetGene (http://

www.cbs.dtu.dk/services/NetGene2/) and the Berkeley Drosophila Genome Project (http://www.fruitfly.org/seq_tools/splice.html).

Mutation screening

PCR primer sets were designed to confirm putative mutations found in *VHL*, *PBRM1*, *BAP1* and *ZNF197*. In addition, 36 primer sets were designed to screen our panel of ccRCC-derived cell lines for mutations in the coding exons and exon- intron boundaries of *VHL* and *PBRM1*. Amplified PCR products were purified using EXOSAP IT (GE Healthcare, Little Chalfont, Buckinghamshire, UK) and sequenced using BigDye Terminator chemistry (Applied Biosystems, Foster City, CA, USA). Reaction products were run on the ABI3130XL Genetic Analyzer (Applied Biosystems). Sequences were evaluated using Mutation Surveyor (SoftGenetics LCC, State college, PA, USA). Primer sequences are available upon request.

RESULTS AND DISCUSSION

SNP analysis

SNP array analysis on 13 ccRCC tumors revealed an allellic imbalance of 3p, most likely indicating copy number loss, in ten tumors. Seven primary ccRCC tumors appeared to have lost the entire short arm of chromosome 3 (3p). Tumor samples T-b, T-c and T-d had copy number loss of part of chromosome 3, with a segment of overlap from pter to 64.0 Mb relative to pter, which includes the 3p21 and 3p25 regions (Supp.Fig. S1).

3p exome sequencing

We conducted partial resequencing of the DNA from the ten primary ccRCC tumors that showed loss of 3p. We enriched for all microRNAs and all exons, exon-intron boundaries, 5'UTRs and 3'UTRs of 542 genes located on the short arm of chromosome 3 and four additional genes (*UTX*, *JARID1C*, *TP53* and *ISW*), representing a total of 2.1 Mb bases. Sequencing resulted in an average of 8.7 million paired-end reads per sample that could be aligned against chromosome 3 (GRCh 37.p2) and to the genomic segments of *UTX*, *JARID1C*, *TP53* and *ISW*. Variant calling and quality filtering, which is described in the Materials and Methods section, resulted in a list of 203 unique putative variations in 136 genes (Supp.Table S1). These include 11 indels introducing a frameshift, 3 inframe indels, 46 missense mutations, 2 mutations that were predicted to affect splicing, 7 variations affecting a 5'UTR, 59 mutations affecting a 3'UTR, 52 intronic mutations with unknown effect and 23 synonymous mutations. (Supp.Table S1). The mutations in *VHL*, *PBRM1*, *BAP1* and *ZNF197* were confirmed using capillary Sanger sequencing (Supp.Fig.S2).

VHL and *PBRM1* are the only genes that show truncating mutations in more than one primary ccRCC tumor sample. Notably, because all the tumor samples showed loss of one copy of 3p, these hemizygous mutations result in complete loss of the functional proteins. We identified five pathogenic mutations in *VHL*. Frameshift introducing mutations in *VHL* were identified in three out of ten primary tumors.

Tumor T-a shows a missense mutation in *VHL* that affects the β- domain and which is predicted to affect protein function and tumor T-j shows an exonic mutation which results in loss of a donor splice site (Supp.Table S1).

Seven out of the ten primary ccRCC tumors proved to have mutations in *PBRM1*. Five of them are truncating mutations. In addition, we identified a missense mutation in *PBRM1* affecting a bromodomain and predicted to be pathogenic in tumor T-h and a splice site mutation resulting in loss of an acceptor site in tumor T-a (Fig. 1, Supp. Table S1).

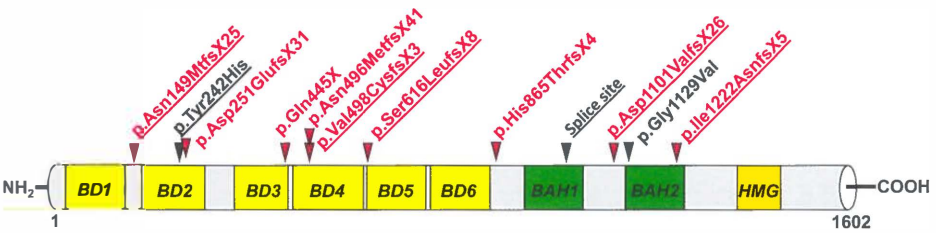


Figure 1. PBRM1 mutations in ccRCC samples. Schematic representation of PBRM1 with arrows showing the locations of the mutations and the corresponding amino acid changes. Red: truncating mutations. Black: missense mutations. Underscored: mutations in primary tumor samples. BD: Bromo domain, BAH: bromo-adjacent homology domain, HMG: high mobility group. Fs: frameshift. National Center for Biotechnology Reference Sequence: NP_060635.2

We did not identify mutations in *SETD2* (3p21), *JARID1C* or *UTX* in this screen, nor did we detect mutations in the tumor suppressors *TP53*, *MLH1*, or other candidate tumor suppressors from 3p including *FHIT*, *TUSC2*, and *FOXP1*.

We previously identified four inactivating mutations in *SETD2* in our panel of ten ccRCC-derived cell lines [Duns et al., 2010]. Capillary Sanger sequencing now revealed inactivating mutations in *PBRM1* and *VHL* in five and six ccRCC cell lines, respectively (Fig. 1, Supp.Fig. S3, Supp.Table S2). Four primary tumor samples and four cell lines had inactivating mutations in both *VHL* and *PBRM1* (Table 1). Three cell lines showed inactivation of *SETD2*, *PBRM1* and *VHL* (Table 1). This suggests that inactivation of these genes affects separate pathways that all contribute to ccRCC tumorigenesis.

VHL regulates the cellular response to oxygen availability in the local microenvironment. Under normoxic circumstances, VHL targets the transcription factor Hypoxia inducible factor α (HIFα) for proteosomal degradation. Inactivation of *VHL* results in consistently high HIFα levels, resulting in upregulation of HIF targets, which include *Vascular endothelial growth factor* (*VEGF*) and *Platelet derived growth factor* (*PDGF*). Binding of these ligands to their cognate receptors activates a series of kinase-dependent signaling pathways including the *RAF-MEK-ERK* and *Phosphatidylinositol-3 kinase-AKT-mTOR* pathways, triggering angiogenesis and proliferation [Clark, 2009].

The functions of *SETD2* and *PBRM1* in tumorigenesis are less well understood. The histone methyltransferase SETD2 is responsible for trimethylation of lysine 36

Table 1. Overview mutations in *VHL*, *PBRM1*, *SETD2*, *BAP1* and *ZNF197* in ccRCC-derived cell lines and primary ccRCC tumors.

Primary ccRCC tumors

Gene	Refseq ID	a	b	c	d	e	f	g	h	i	j
PBRM1	NM_018165.4	Splice site			ins 1 bp	del 1 bp	ins 1 bp	del 1 bp	missense		del 1 bp
VHL	NM_000551.2	missense	del 2 bp			ins 1 bp		del 4 bp			Splice site
SETD2	NM_014159.6										
BAP1	NM_004656.2		del 1 bp								
ZNF197	NM_006991.3						del 4 bp				

ccRCC derived cell lines

Gene	Refseq ID	1	4	5	AB	ER	FG2	HS	JF	JW	MF
PBRM1	NM_018165.4		ins 1 bp			del 4 bp	Nonsense	missense			del 1 bp
VHL	NM_000551.2	ins 37 bp	missense			del 1 bp	del 1 bp			ins 6 bp	missense
SETD2	NM_014159.6				del 1 bp	Nonsense	Loss 3 exons			missense	del 9 bp

ins:insertion, del: deletion, bp: basepair(s)

on histone 3 (H3K36me3), a modification which is associated with transcriptionally active genes and which has been suggested to play a role in transcriptional elongation and the regulation of alternative splicing of target genes [Luco et al., 2010]. *SETD2* inactivation appears to be unique for ccRCC tumors, suggesting a tumor suppressive function of SETD2 specifically in proximal tubular epithelial cells [Dalgliesh et al., 2010; Duns et al., 2010]. In this screen, we did not identify inactivating mutations in *SETD2*.

PBRM1 is located at 3p21 and encodes the Polybromo 1 (BAF180) protein, which is the chromatin targeting subunit of the Polybromo complex SWI/SNF [Thompson, 2009]. SWI/SNF complexes are large ATP-dependent chromatin-remodeling machines that change the accessibility of DNA by controlling the dynamics of nucleosome occupancy [Reisman et al., 2009]. Several SWI/SNF components have previously been shown to be involved in cancer development: inactivating mutations in *SMARCB1* and *ARID1A* have been identified in rhabdoid tumors and ovarian carcinomas respectively [Reisman et al., 2009; Jones et al., 2010]. Inactivating mutations in *BRG1* have been identified in several tumor types [Rodriguez-Nieto et al., 2011; Schneppenheim et al., 2010], while truncating mutations in *PBRM1* have been described previously in breast cancer samples [Xia et al., 2008].

The role of SWI/SNF complexes in tumor suppression is poorly understood. SWI/SNF components target several genes and signaling pathways that are involved in cell proliferation and cancer, including the Wnt, ErbB, p53, MAPK and insulin signaling pathways, and in processes such as DNA repair, cell cycle regulation and apoptosis [Euskirchen et al., 2011]. PBRM1 plays a role in coordinating senescence by regulating the transcriptional activity of a subset of p53 target genes [Burrows et al., 2010], and has been suggested to be involved in the cellular response to hypoxia [Kenneth et al., 2009].

The identification of inactivating mutations in *PBRM1* in primary ccRCC tumors and ccRCC-derived cell lines confirms the recent findings of *Varela et al*, who sequenced the whole exome of seven primary ccRCC tumors and corresponding normal tissue, and identified somatic truncating mutations in *PBRM1* in four of them [Varela et al., 2011]. Mutational screening of *PBRM1* using capillary-based sequencing in a total of 227 primary ccRCC tumors revealed inactivating mutations in 91 (41 %) of them. The frequency and nature of the mutations in *PBRM1* in ccRCC tumors suggest that it is a major ccRCC tumor suppressor gene. This points to an important role for aberrant chromatin regulation in the development of ccRCC and proves the longstanding hypothesis that 3p21 is the location of a major tumor suppressor gene.

We identified truncating mutations in two additional genes: *ZNF197* and *BAP1* (Table 1, Supp.Table S1). *BAP1*, located at 3p21.31 and coding for BRCA1 associated protein-1, is especially interesting with respect to ccRCC pathogenesis. BAP1 mediates deubiquitination of histone H2A and HCHC1, and is involved in the transcriptional control of genes regulating cell growth and proliferation [Yu et al., 2010]. Frequent mutations in *BAP1* have been identified in metastizing uveal melanomas and malignant pleural mesothelioma [Harbour et al., 2010; Bott et al., 2011]. A role for *BAP1* inactivation in a subset of ccRCC tumors would be consistent with the emerging role of aberrant chromatin regulation in ccRCC. *BAP1* was not included in the screen

by *Dalgliesh et al*, and no inactivating mutations were identified in the seven ccRCC tumors sequenced by *Varela et al* [Dalgliesh et al., 2010; Varela et al., 2011]. Mutational screening of *BAP1* in a larger panel of ccRCC tumors should therefore give a better estimate of the frequency of inactivation in ccRCC.

ZNF197 belongs to the zinc finger protein superfamily, members of which are regulatory proteins characterized by nucleic acid-binding zinc finger domains. Alternative isoform 2 of *ZNF197* lacks these zinc finger motifs and codes for a protein that has been shown to bind to pVHL and negatively regulate HIF-1-α transactivation [Li et al., 2003]. Inactivation of *ZNF197* might therefore be redundant to *VHL* inactivation. The mutation in *ZNF197* was identified in a sample in which no *VHL* mutation was identified (Table 1).

ccRCC genetics

ccRCC genetics appears to be dominated by inactivation of two tumor suppressor genes, *VHL* and *PBRM1*, as the majority of primary ccRCC tumors and ccRCC derived cell lines have inactivating mutations in *VHL* and/or *PBRM1*. Other genes, including *SETD2*, *JARID1C* and *UTX*, appear to play a role in a smaller subset of ccRCC tumors [Dalgliesh et al., 2010]. We did not identify inactivating mutations in these genes in our panel of primary ccRCC tumors, nor did we detect any mutations in several other (candidate) tumor suppressor genes. Several of our ccRCC derived cell lines show concomitant inactivating mutations in *VHL*, *PBRM1* and *SETD2*, suggesting that inactivation of these genes is functionally nonredundant. Further research is needed to define the specific roles of *VHL*, *PBRM1* and *SETD2* in ccRCC tumorigenesis. Our ccRCC-derived cell lines, with their known mutational status for these 3 genes, are ideal tools for performing further research.

ACKNOWLEDGEMENTS

This work was funded by the Dutch Cancer Society, grant 2007-3892.
We thank Jackie Senior for editing the paper.

REFERENCES

Bott M, Brevet M, Taylor BS, et al. The nuclear deubuquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. Nature Genet 2011;43:668.

Burrows AE, Smogorzewska A, Elledge SJ. Polybromo-associated BRG-associated factor components BRD7 and BAF180 are critical regulators of p53 required for induction of replicative senescence. PNAS U S A 2010;107:14280.

Clark PE. The role of VHL in clear-cell renal cell carcinoma and its relation to targeted therapy. Kidney Int 2009;76:939-945.

Dalgliesh GL, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying enzymes. Nature 2010;463:360.

Duns G, van den Berg E, van Duivenbode I, et al. Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. Cancer Res 2010;70:4287.

Euskirchen GM, Auerbach RK, Davidov E, et al. Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. PloS Genet 2011;7:e1002008.

Harbour JW, Onken MD, Roberson ED, et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. Science 2010;330:1410.

Jones S, Wang TL, Shih leM, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science 2010;330:228.

Kenneth NS, Mudie S, Van Uden P, et al. SWI/SNF regulates the cellular response to hypoxia. J Biol Chem 2009;284:4123.

Li Z, Wang D, Na X, Schoen SR, et al. The VHL protein recruits a novel KRAB-A domain protein to repress HIF-1alpha transcriptional activity. EMBO J 2003;22:1857.

Luco RF, Pan Q, Tominaga K, et al. Regulation of alternative splicing by histone modifications. Science 2010;327:996.

Mandriota SJ, Turner KJ, Davies DR, et al. HIF inactivation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron. Cancer Cell 2002;1:459.

Reisman D, Glaros S, Thompson EA. The SWI/SNF complex and cancer. Oncogene 2009;28:1653.

Rodriguez-Nieto S, Canada A, Pros E, et al. Massive parallel DNA pyrosequencing analysis of the tumor suppressor BRG1/SMARCA4 in lung primary tumors. Hum Mutat 2011;32: 1999.

Schneppenheim R, Frühwald MC, Gesk S, et al. Germline mutation and somatic inactivation of *SMARCA4/BRG1* in a family with rhabdoid tumor predisposition syndrome. Am J Hum Genet 2010;86:279.

Thompson M. Polybromo-1: the chromatin targeting subunit of the PBAF complex. Biochimie 2009;91:309.

Van den Berg A, Dijkhuizen T, Draaijers TG, et al. Analysis of multiple renal cell adenomas and carcinomas suggests allelic loss at 3p21 to be a prerequisite for malignant development. Genes Chromosomes Cancer 1997;19:228.

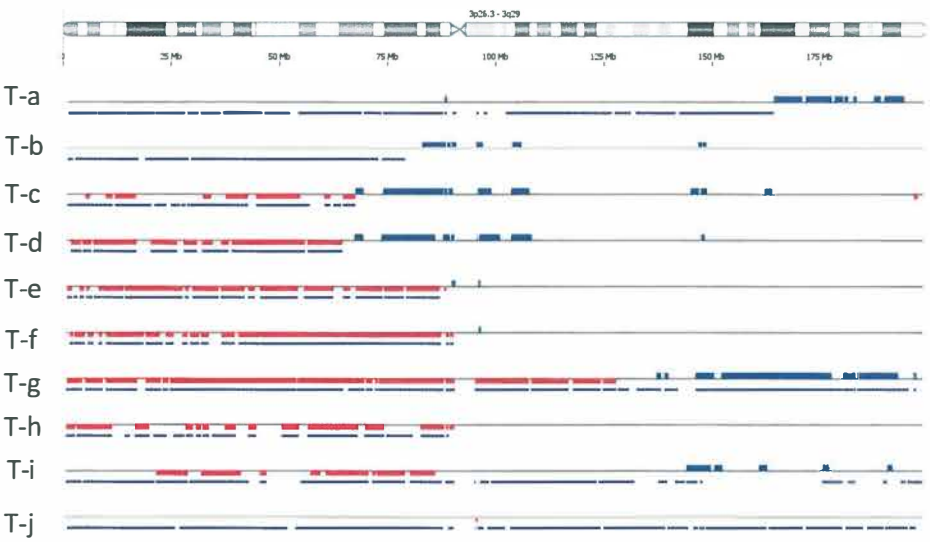
Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 2011;469:539.

Xia W, Nagase S, Montia AG, et al. BAF180 is a critical regulator of p21 induction and a tumor suppressor mutated in breast cancer. Cancer Res 2008;68:1667'.

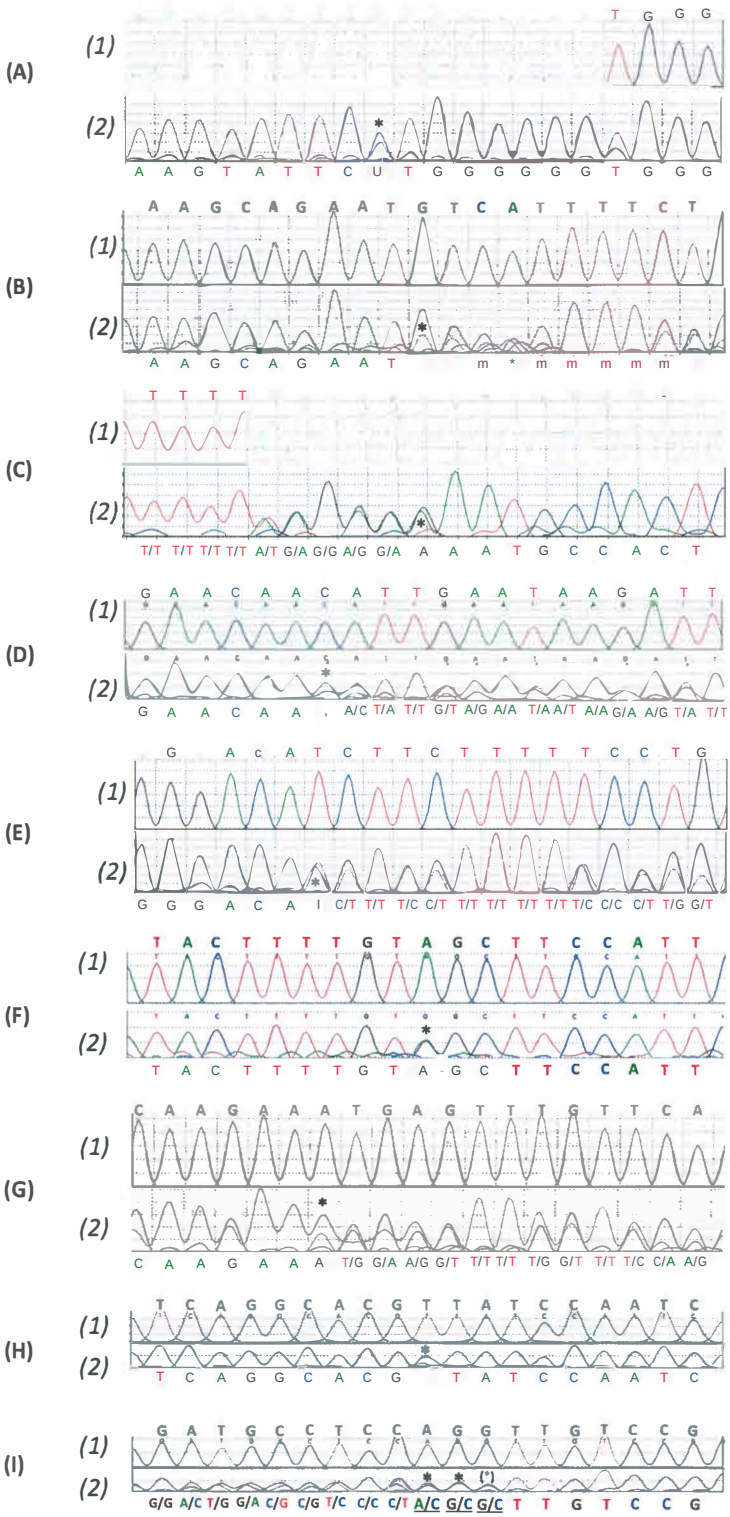
Young AP, Schlisio S, Minamishima YA, et al. VHL loss actuates a HIF- independent senescence programme mediated by Rb and p400. Nat Cell Biol 2008;10:361.

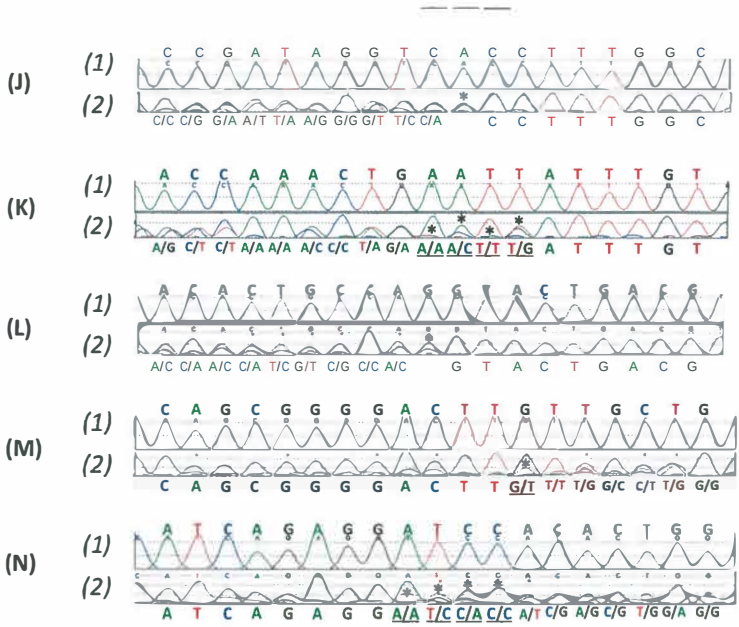
Yu H, Mashtalir N, Daou S, et al. The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. Mol Cell Biol 2010;30:5071.

SUPPLEMENTARY FIGURES AND TABLES

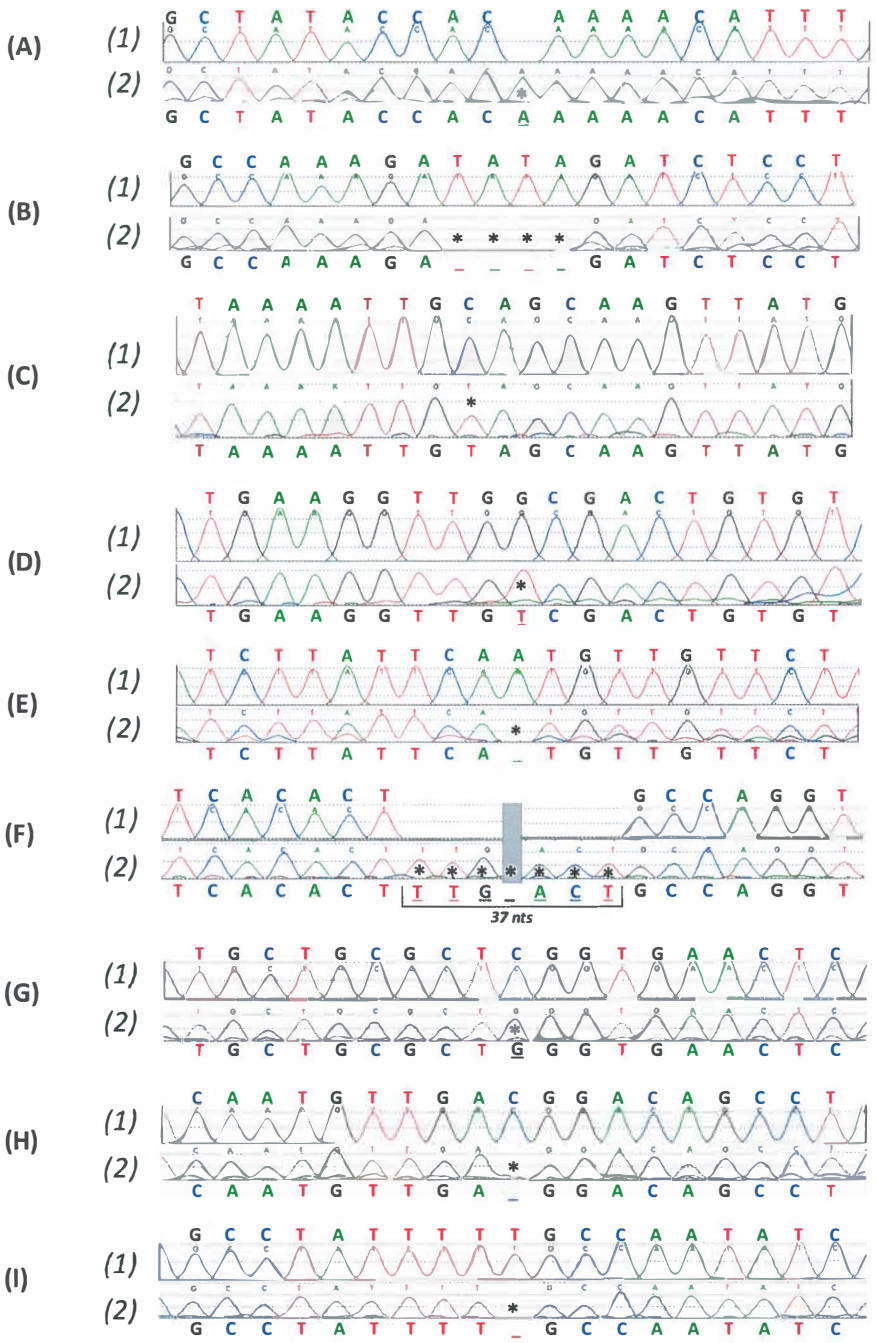


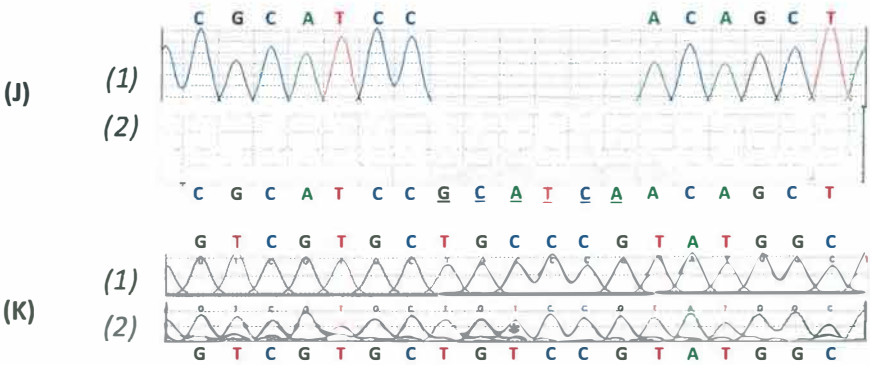
Supplementary Figure 1. 3p copy number loss in primary ccRCC tumors assessed by SNP arrays. Red lines: regions showing chromosomal loss. Blue horizontal lines: regions showing chromosomal gain. Purple line: regions showing imbalance.


(A)
(1)
(2)
(B)
(1)
(2)
(C)
(1)
(2)
(D)
(1)
(2)
(E)
(1)
(2)
(F)
(1)
(2)
(G)
(1)
(2)
(H)
(1)
(2)
(I)
(1)
(2)



Supplementary Figure 2. Confirmation mutations identified by exome sequencing in *PBRM1* (a-g), *VHL* (h-l), *BAP1* (m) and *ZNF197* (n) by capillary Sanger sequencing. A. Sequence analysis of part of *PBRM1*. (1) Wild type. (2) Tumor T-a, showing an intronic mutation affecting a splice site (NM_018165.4:c.3082-1G>A) (*). **B.** Sequence analysis of part of *PBRM1*. (1) Wild type. (2) Tumor T-d, showing a 1 bp insertion introducing a frameshift resulting in a premature termination codon (NM_018165.4:c3664dup, p.Ile1222AsnfsX5) (*) **C.** Sequence analysis of part of *PBRM1*. (1) Wild type. (2) Tumor T-e, showing a 1 bp deletion introducing a frameshift resulting in a premature termination codon (NM_018165.4:c1846del, p.Ser616LeufsX8) (*) **D.** Sequence analysis of part of *PBRM1*. (1) Wild type. (2) Tumor T-f, showing a 1 bp insertion introducing a frameshift resulting in a premature termination codon (NM_018165.4: c.1491dup, p.Val498CysfsX3) (*). **E.** Sequence analysis of part of *PBRM1*. (1) Wild type. (2) Tumor T-g, showing a 1 bp deletion introducing a frameshift resulting in a premature termination codon (NM_018165.4:c3302del, p.Asp1101ValfsX26) (*). **F.** Sequence analysis of part of *PBRM1*. (1) Wild type. (2) Tumor T-h, showing a missense mutation (NM_018165.4: c.724T>C, pTyr242His) (*). **G.** Sequence analysis of part of *PBRM1*. (1) Wild type. (2) Tumor T-j, showing a 1 bp deletion introducing a frameshift resulting in a premature termination codon (NM_018165.4:c.446del, p.Asn149MetfsX25) (*). The observed additional upstream frameshift is the result of an intronic 1 bp deletion. **H.** Sequence analysis of part of *VHL*. (1) Wild type. (2) Tumor T-a, showing a missense mutation (NM_000551.2:c.349T>G,p.Trp117Gly) (*). **I.** Sequence analysis of part of *VHL*. (1) Wild type. (2) Tumor T-b, showing a 2 bp deletion introducing a frameshift resulting in a premature termination codon (NM_000551.2:c.491-492del, p.Gln164ArgfsX9) (*). Additional mutation: c.493G>C. **J.** Sequence analysis of part of *VHL*. (1) Wild type. (2) Tumor T-e, showing a 1 bp insertion introducing a frameshift resulting in a premature termination codon (NM_000551.2:c.346dup, p.Leu116ProfsX16) (*). **K.** Sequence analysis of part of *VHL*. (1) Wild type. (2) Tumor T-g, showing a 4 bp deletion introducing a frameshift resulting in a premature termination codon (NM_000551.2:c401-404del, p.Glu134AspfsX24) (*). **L.** Sequence analysis of part of *VHL*. (1) Wild type. (2) Tumor T-j, showing a 1 bp deletion affecting a splice site (NM_000551.2:c.463+1del) (*) **M.** Sequence analysis of part of *BAP1*. (1) Wild type. (2) Tumor T-b, showing a 1 bp insertion introducing a frameshift resulting in a premature termination codon (NM_0004656.2:c870del, p.Asn290LysfsX45) (*) **N.** Sequence analysis of part of *ZNF197*. (1) Wild type. (2) Tumor T-f, showing a 4 bp deletion introducing a frameshift resulting in a premature termination codon (NM_006991.3:c.1339-1342del, p.Ile447ThrfsX22) (*).


(A)
(1)
(2)
G C T A T A C C A C A A A A A C A T T T
G C T A T A C C A C A A A A A C A T T T
(B)
(1)
(2)
G C C A A A G A T A T A G A T C T C C T
G C C A A A G A _ _ _ _ G A T C T C C T
(C)
(1)
(2)
T A A A A T T G C A G C A A G T T A T G
T A A A A T T G T A G C A A G T T A T G
(D)
(1)
(2)
T G A A G G T T G G C G A C T G T G T
T G A A G G T T G T C G A C T G T G T
(E)
(1)
(2)
T C T T A T T C A A T G T T G T T C T
T C T T A T T C A _ T G T T G T T C T
(F)
(1)
(2)
T C A C A C T G C C A G G T
T C A C A C T T T G _ A C T G C C A G G T
37 nts
(G)
(1)
(2)
T G C T G C G C T C G G T G A A C T C
T G C T G C G C T G G G T G A A C T C
(H)
(1)
(2)
C A A T G T T G A C G G A C A G C C T
C A A T G T T G A _ G G A C A G C C T
(I)
(1)
(2)
G C C T A T T T T T G C C A A T A T C
G C C T A T T T T _ G C C A A T A T C



Supplementary Figure 3. Mutations in *PBRM1* (a-e) and *VHL* (f-k) in ccRCC derived cell lines identified by capillary Sanger sequencing. A. Sequence analysis of *PBRM1*. (1) Wild type. (2). RCC-4, showing a 1 bp insertion introducing a frameshift resulting in a premature termination codon (NM_018165.4:c2592dup, p.His865ThrfsX4) (*). **B.** Sequence analysis of *PBRM1*. (1) Wild type. (2). RCC-ER, showing a 4 bp deletion introducing a frameshift resulting in a premature termination codon (NM_018165.4:c753-756del, p.Asp251GlufsX31) (*). **C.** Sequence analysis of *PBRM1*. (1) Wild type. (2) RCC-FG2, showing a nonsense mutation (NM_018165.4:c1333C>T, pGln445X) (*). **D.** Sequence analysis of *PBRM1*. (1) Wild type. (2) RCC-HS, showing a missense mutation (NM_018165.4: c.3386G>T, pGly1129Val) (*). **E.** Sequence analysis of *PBRM1*. (1) Wild type. (2) RCC-MF, showing a 1 bp deletion introducing a frameshift resulting in a premature termination codon (NM_018165.4:c1487del, p.Asn496MetfsX4)(*). **F.** Sequence analysis of *VHL*. (1) Wild type. (2) RCC-1, showing a 37 bp insertion introducing a frameshift resulting in a premature termination codon (NM_000551.2:c458-459ins37, p.Pro154CysfsX2) (*) **G.** Sequence analysis of *VHL*. (1) Wild type. (2) RCC-4, showing a missense mutation (NM_000551.2: c.194C>G, pSer65Trp) **H.** Sequence analysis of *VHL*. (1) Wild type. (2) RCC-ER, showing a 1 bp deletion introducing a frameshift resulting in a premature termination codon (NM_000551.2:c482del, p.Asp143AlafsX16) (*). **I.** Sequence analysis of *VHL*. (1) Wild type. (2) RCC-FG2, showing a 1 bp deletion introducing a frameshift resulting in a premature termination codon (NM_000551.2:c.444del, p.Phe148LeufsX11) (*). **J.** Sequence analysis of *VHL*. (1) Wild type. (2) RCC-JW, showing an inframe 6 bp insertion (NM_000551.2:c.328-444insGCATCA, p.His110delinsArgIleAsn) (*). **K.** Sequence analysis of *VHL*. (1) Wild type. (2) RCC-MF, showing a missense mutation (NM_000551.2:c.236C>T, p.Pro86Ser) (*).

Supplementary Table S1. Unique variations identified by targeted exome sequencing.

				Mutation		
Gene name	Refseq	Tumor	Sanger seq	type	genomic location	cDNA level
ABHD5	NM_016006	T-a	not performed	missense	43743914	NM_016006.4 (ABHD5):c.341G>T
ABHD5	NM_016006	T-c	not performed	missense	43743774	NM_016006.4 (ABHD5):c.201G>C
ADAMTS9	NM_182920	T-c	not performed	3'UTR	64501719	
APEH	NM_001640	T-a	not performed	intronic	49715123	
APPL1,ASB14	NM_012096	T-b	not performed	3'UTR	57303851	
APPL1,ASB14	NM_001142733	T-b	not performed	intronic	57303851	
ARHGEF3	NM_001128615	T-g	not performed	intronic	56779170	
ARHGEF3	NM_019555.2	T-i	not performed	missense	56787540	NM_019555.2 (ARHGEF3):c.430G>T
ARIH2	NM_006321	T-b	not performed	intronic	49004749	
ARPP21	NM_001025068	T-g	not performed	3'UTR	35726817	
ARPP21	NM_016300	T-g	not performed	intronic	35726817	
ATG7	NM_001144912.1	T-e	not performed	missense	11383650	NM_001144912.1 (ATG7):c.905G>T
ATXN7	NM_000333	T-b	not performed	3'UTR	63985157	
ATXN7	NM_000333	T-g	not performed	3'UTR	63986514	
ATXN7	NM_001128149	T-b	not performed	missense	63985157	NM_001177387.1 (ATXN7):c.2760C>A
BAP1	NM_004656.2	T-b	Confirmed	Frameshift	52439842	NM_004656.2 (BAP1):c.870del
BAP1	NM_004656.2	T-j	not performed	intronic	52436770	
BSN	NM_003458	T-a	not performed	missense	49691434	NM_003458.3 (BSN):c.4445C>T
C3orf18	NM_016210	T-b	not performed	3'UTR	50596371	
C3orf77	NM_001145030	T-h	not performed	synonymous	44328924	
CACNA2D2	NM_001005505	T-c	not performed	3'UTR	50401376	
CACNA2D2	NM_001005505	T-e	not performed	3'UTR	50400737	
CACNA2D2	NM_001005505	T-i	not performed	3'UTR	50401336	
CACNA2D3	NM_018398	T-i	not performed	intronic	54354624	
CACNA2D3	NM_018398	T-h	not performed	synonymous	55107571	
CADM2	NM_153184	T-d	not performed	3'UTR	86117618	
CADM2	NM_153184	T-d	not performed	synonymous	85935413	
CAPN7	NM_014296	T-i	not performed	missense	15259048	NM_014296.2 (CAPN7): c.328A>G
CAV3	NM_001234	T-i	not performed	5'UTR	8775526	
CAV3	NM_033337.2	T-h	not performed	missense	8787347	NM_033337.2 (CAV3): c.250C>G
CCDC36	NM_178173.3	T-h	Not confirmed	missense	49293819	NM_178173.3 (CCDC36): c889T>G
CCDC66	NM_001012506	T-j	not performed	intronic	56651698	
CCDC66	NM_001012506.4	T-g	not performed	missense	56653900	NM_001012506.4 (CCDC66): c.2629A>G
CCR3	NM_001837	T-a	not performed	5'UTR	46283902	

				Frequency	Predictions		
protein level	Coverage	Variant	WT	Variant/WT	AGVGD	SIFT	Polyphen
p.Arg114Leu	787	T	G	34.9/65.1	C0	Tolerated	Probably damaging
Lys67Asn	348	C	G	24.7/75.3	C0	Tolerated	Benign
x	165	T	C	44.2/55.8	x	x	x
	255	A	G	62.4/37.6	x	x	x
x	281	T	A	32.0/68.0	x	x	x
	281	T	A	68.0/32.0	x	x	x
	219	A	C	75.3/24.7	x	x	x
Ala144Ser	216	A	C	62.0/38.0	C65	Affect protein function	Possibly damaging
	196	A	G	70.4/29.6	x	x	x
x	335	G	A	78.8/21.2	x	x	x
	335	G	A	78.8/21.2	x	x	x
Cys302Phe	259	T	G	44.8/55.2	C0	Tolerated	Possibly damaging
x	474	A	C	22.6/77.4	x	x	x
x	311	C	A	31.8/68.2	x	x	x
Ser920Arg	474	A	C	77.4/22.6	C0	Tolerated	Possibly damaging
p.Asn290LysfsX45	192	-	G	36.5/63.5	x	x	x
	306	A	T	34.3/65.7	x	x	x
p.Pro1482Leu	426	T	C	54.5/45.5	C65	Affect protein function	Probably damaging
x	432	T	C	25.2/74.8	x	x	x
x	221	C	A	67.3/32.7	x	x	x
x	237	T	C	71.3/28.7	x	x	x
x	336	A	G	43.8/56.2	x	x	x
x	242	C	G	29.8/70.2	x	x	x
	177	T	C	32.8/67.2	x	x	x
x	284	T	C	63.7/36.3	x	x	x
x	261	T	C	19.9/81.1	x	x	x
x	365	T	C	71.5/28.5	x	x	x
Ile110Val	398	G	A	61.1/38.9	C0	Tolerated	Benign
x	281	A	G	28.1/71.9	x	x	x
p.Leu84Val	338	G	C	63.0/37.0	C0	Tolerated	Benign
Trp297Gly,Gly	249	G	T	73.1/26.9	C0	Tolerated	Probably damaging
	157	A	G	41.4/58.6	x	x	x
Ile877Val	616	G	A	73.7/26.3	C0	Tolerated	Probably damaging
x	703	T	C	37.7/62.3	x	x	x

Supplementary Table S1. Continued.

Gene name	Refseq	Tumor	Sanger seq	Mutation type	genomic location	cDNA level
CCR5	NM_000579	T-d	not performed	3'UTR	46417340	
CCR5	NM_000579	T-g	not performed	3'UTR	46417266	
CDCP1	NM_022842	T-c	not performed	missense	45130627	NM_022842.3 (CDCP1): c.2000C>G
CELSR3	NM_001407	T-g	not performed	3'UTR	48675391	
CGGBP1	NM_001008390	T-e	not performed	3'UTR	88101730	
CGGBP1	NM_001008390	T-g	not performed	3'UTR	88101339	
CHDH	NM_018397	T-j	not performed	3'UTR	53851505	
CHDH	NM_018397	T-g	not performed	intronic	53852920	
CLASP2	NM_015097	T-f	not performed	intronic	33653588	
CMTM8	NM_178868.3	T-g	not performed	missense	32398997	NM_178868.3 (CMTM8): c.280A>G
CNTN3	NM_020872	T-b	not performed	missense	74413659	NM_020872.1 (CNTN3): c.1172A>T
CNTN4	NM_175612	T-d	not performed	5'UTR	3081519	
CNTN4	NM_175607	T-i	not performed	intronic	3097734	
CNTN4	NM_175607	T-g	not performed	synonymous	3081933	
CNTN6	NM_014461	T-i	not performed	synonymous	1337371	
COL7A1	NM_000094	T-a	not performed	intronic	48620392	
COLQ	NM_005677	T-d	not performed	3'UTR	15491818	
CPNE9	NM_153635	T-j	not performed	missense	9756618	NM_153635.2 (CPNE9): c.671A>G
CTNNB1	NM_001098209	T-f	not performed	intronic	41268866	
CTNNB1	NM_001904.3	T-b	not performed	missense	41266697	NM_001904.3 (CTNNB1): c.494A>G
DAG1	NM_004393	T-b	not performed	3'UTR	49572438	
DAG1	NM_004393	T-d	not performed	3'UTR	49572260	
DCP1A	NM_018403	T-g	not performed	missense	53378932	NM_018403.5 (DCP1A): c.173G>A
DHFRL1	NM_176815	T-f	not performed	missense	93780283	NM_001195643.1 (DHFRL1): c.73A>T
DLEC1	NM_007337	T-b	not performed	in frame del	38164071	NM_007337.2 (DLEC1): c.5312-5314del
DLEC1	NM_007337	T-i	not performed	missense	38163900	NM_007337.2 (DLEC1): c.5141C>T
DLEC1	NM_007335	T-i	not performed	synonymous	38163900	
DNAH1	NM_015512	T-b	not performed	missense	52404178	NM_015512.4 (DNAH1): c.6191A>G
DNAH1	NM_015512	T-d	not performed	missense	52420315	NM_015512.4 (DNAH1): c.8765C>T
DNAH1	NM_015512	T-b	not performed	synonymous	52386615	
DOCK3	NM_004947	T-e	not performed	intronic	51112909	
DOCK3	NM_004947	T-b	not performed	synonymous	51399347	
ENTPD3	NM_001248	T-e	not performed	missense	40442421	NM_001248.2 (ENTPD3): c.205A>G

				Frequency	Predictions		
protein level	Coverage	Variant	WT	Variant/WT	AGVGD	SIFT	Polyphen
x	670	T	A	32.2/67.8	x	x	x
x	630	T	G	22.4/77.6	x	x	x
Thr667Ser	237	C	G	39.7/60.3	C0	Tolerated	Benign
x	328	C	G	72.0/28.0	x	x	x
x	714	A	C	40.2/59.8	x	x	x
x	553	C	T	49.9/50.1	x	x	x
x	307	T	C	64.2/35.8	x	x	x
	170	A	G	60.0/40.0	x	x	x
	284	C	T	64.8/35.2	x	x	x
Met94Val	363	G	A	70.5/29.5	C0	Tolerated	Benign
Lys391Ile	370	A	T	43.0/57.0	C0	Affect protein function	Benign
x	481	G	A	25.8/74.2	x	x	x
	184	T	G	63.0/37.0	x	x	x
x	367	A	G	21.0/79.0	x	x	x
x	597	C	T	27.0/73.0	x	x	x
	171	A	C	55.0/45.0	x	x	x
x	412	A	G	30.1/69.9	x	x	x
Tyr224Cys	283	G	A	18.7/81.3	C65	Affect protein function	Probably damaging
	162	G	A	60.5/38.9	x	x	x
Gln165Arg	417	G	A	64.7/35.3	C0	Tolerated	Possibly damaging
x	211	A	G	26.1/73.9	x	x	x
x	665	C	G	62.9/37.1	x	x	x
Arg58Gln	293	T	C	47.1/52.9	C0	Tolerated	Probably damaging
Arg25Trp	921	A	T	40.7/59.3	C0	Tolerated	Benign
p.Asp1772del	253	---	ATG	58.1/41.9	x	x	x
Pro1714Leu	246	T	C	64.6/35.4	C0	Tolerated	Benign
x	246	C	T	35.4/65.6	x	x	x
Asn2064Ser	610	G	A	67.7/32.3	C0	Affect protein function	Benign
Ala2922Val	195	T	C	64.1/35.9	C0	Affect protein function	Benign
x	295	T	C	28.5/71.5	x	x	x
	306	T	C	75.2/24.8	x	x	x
x	646	A	G	31.0/68.9	x	x	x
Thr69Ala	210	G	A	21.4/78.6	C0	Affect protein function	Probably damaging

Supplementary Table S1. Continued.

Gene name	Refseq	Tumor	Sanger seq	Mutation type	genomic location	cDNA level
EPHA3	NM_005233	T-b	not performed	intronic	89445194	
EPM2AIP1	NM_014805	T-i	not performed	3'UTR	37030382	
EPM2AIP1	NM_014805	T-h	not performed	missense	37033820	NM_014805.3 (EPM2AIP1): c.749T>C
ERC2	NM_015576	T-g	not performed	intronic	56041371	
FAM116A	NM_152678	T-b	not performed	missense	57614059	NM_152678.2 (FAM115A): c.1709G>C
FBLN2	NM_001004019	T-g	not performed	synonymous	13672866	
FGD5	NM_152536	T-e	not performed	3'UTR	14974910	
FGD5	NM_152536	T-g	not performed	3'UTR	14862732	
FGD5	NM_152536	T-b	not performed	intronic	14941867	
FOXP1	NM_032682	T-i	not performed	intronic	71161817	
FRMD4B	NM_015123	T-f	not performed	3'UTR	69220061	
FYCO1	NM_024513	T-e	not performed	missense	46008983	NM_024513.2 (FYCO1): c.1843C>T
GADL1	NM_207359	T-j	not performed	intronic	30898466	
GBE1	NM_000158	T-d	not performed	3'UTR	81539250	
GPD1L	NM_015141	T-b	not performed	3'UTR	32207534	
GRM2	XM_002342357	T-b	not performed	5'UTR	51743411	
HACL1	NM_012260	T-d	not performed	synonymous	15609395	
HYAL1	NM_007312	T-i	not performed	5'UTR	50340517	
IFRD2	NM_006764	T-i	not performed	missense	50327685	NM_006764.4 (IFRD2): c.496G>A
IL17RB	NM_018725	T-e	not performed	intronic	53891058	
IL17RC	NM_032732	T-i	not performed	intronic	9972688	
IL17RE	NM_153480	T-h	not performed	missense	9948480	NM_001193380.1 (IL17RE): c.457A>G
IQCF2	NM_203424	T-e	not performed	missense	51897361	NM_203424.1 (IQCF2): c.470T>C
IQSEC1	NM_014869	T-d	not performed	intronic	12941917	
ITIH1	NM_002215	T-b	not performed	synonymous	52817014	
ITIH3	NM_002217	T-a	not performed	intronic	52842507	
ITPR1	NM_001099952	T-a	not performed	intronic	4738980	
KAT2B	NM_003884	T-c	not performed	missense	20189472	NM_003884.4 (KAT2B): c.2137C>A
KBTBD8	NM_032505	T-a	not performed	3'UTR	67060743	
KCNH8	NM_144633	T-e	not performed	intronic	19559422	
KIF15	NM_020242	T-a	not performed	intronic	44869851	
LMOD3	NM_198271.3	T-b	not performed	in frame del	69171497	NM_198271.3 (LMOD3): c.41-43del
LOC100132526	XM_001721212	T-a	not performed	3'UTR	13977985	
LOC401052	NM_001008737	T-h	not performed	3'UTR	10048939	
LRIG1	NM_015541	T-d	not performed	intronic	66449347	
LRIG1	NM_015541	T-d	not performed	intronic	66449347	

				Frequency	Predictions		
protein level	Coverage	Variant	WT	Variant/WT	AGVGD	SIFT	Polyphen
	243	A	G	21.8/78.2	x	x	x
x	152	G	C	35.5/64.5	x	x	x
Leu250Ser	484	G	A	73.8/26.2	x	x	Benign
	253	C	T	32.4/67.6	x	x	x
Arg570Pro	288	G	C	64.2/35.8	C0	Tolerated	Possibly damaging
x	193	T	C	30.6/69.4	x	x	x
x	190	G	C	29.5/70.5	x	x	x
x	383	T	C	21.9/78.1	x	x	x
	262	A	G	57.6/42.2	x	x	x
	220	T	C	54.1/45.9	x	x	x
x	336	A	G	62.5/37.5	x	x	x
Arg615Trp	194	A	G	79.4/20.6	C0	Affect protein function	Probably damaging
	207	C	T	61.4/38.6	x	x	x
x	211	A	G	22.7/77.3	x	x	x
x	447	C	T	17.0/83.0	x	x	x
x	450	A	G	25.8/74.2	x	x	x
x	668	A	G	27.1/72.9	x	x	x
x	365	T	G	63.3/36.7	x	x	x
Ala166Thr	189	T	C	68.8/31.2	C0	Tolerated	Probably damaging
	277	C	T	77.3/22.7	x	x	x
	345	G	C	29.6/70.1	x	x	x
Lys153Glu,Glu	230	G	A	30.0/70.0	C0	Affect protein function	Probably damaging
Phe157Ser	572	C	T	17.3/82.5	C65	Affect protein function	Probably damaging
	277	A	G	77.3/22.7	x	x	x
x	281	G	A	68.3/31.7	x	x	x
	177	T	C	70.1/29.9	x	x	x
	235	A	G	44.7/55.3	x	x	x
Pro713Thr	221	A	C	20.8/79.2	C35	Affect protein function	Probably damaging
x	435	C	A	34.9/65.1	x	x	x
	205	G	A	61.0/39.0	x	x	x
	158	G	A	63.3/37.7	x	x	x
p.Leu14_Asp15delinsHis	358	---	AGA	27.4/72.6	x	x	x
x	235	G	C	57.9/42.1	x	x	x
x	188	G	A	67.0/33.0	x	x	x
	183	T	C	30.1/69.9	x	x	x
	183	T	C	30.1/69.9	x	x	x

Supplementary Table S1. Continued.

				Mutation		
Gene name	Refseq	Tumor	Sanger seq	type	genomic location	cDNA level
LRRC3B	NM_052953	T-g	not performed	3'UTR	26751986	
LTF	NM_002343	T-f	not performed	synonymous	46496910	
MAGI1	NM_004742	T-g	not performed	3'UTR	65346411	
MAGI1	NM_001033057	T-i	not performed	3'UTR	65341301	
MAGI1	NM_004742	T-i	not performed	3'UTR	65345591	
MAGI1	NM_001033057	T-f	not performed	intronic	65345134	
MAGI1	NM_001033057	T-g	not performed	intronic	65346411	
MAGI1	NM_001033057	T-i	not performed	intronic	65345591	
MAGI1	NM_001033057	T-e	not performed	missense	65456135	NM_015520.1 (MAGI1): c.782A>G
MAGI1	NM_002343	T-f	not performed	synonymous	65345134	
MANF	NM_006010	T-g	not performed	intronic	51425141	
MAP4	NM_001134364	T-g	not performed	3'UTR	47892855	
MITF	NM_000248	T-j	not performed	5'UTR	69985764	
NEK10	NM_199347	T-d	not performed	intronic	27387564	
NEK10	NM_199347	T-b	not performed	missense	27338698	NM_199347.2 (NEK10): c.1202A>G
NISCH	NM_007184	T-c	not performed	intronic	52504963	
NISCH	NM_007184	T-c	not performed	synonymous	52504895	
NKIRAS1	NM_020345	T-h	not performed	3'UTR	23934322	
NR1D2	NM_005126.4	T-e	not performed	missense	24003660	NM_005126.4 (NR1D2): c.710A>G
NR2C2	NM_003298	T-a	not performed	3'UTR	15086206	
NR2C2	NM_003298	T-d	not performed	3'UTR	15085906	
NR2C2	NM_003298	T-g	not performed	3'UTR	15089098	
NR2C2	NM_003298	T-j	not performed	3'UTR	15088874	
NR2C2	NM_003298	T-h	not performed	intronic	15065790	
NUP210	NM_024923	T-a	not performed	3'UTR	13357935	
NUP210	NM_024923	T-b	not performed	missense	13383322	NM_024923.2 (NUP210): c.3154G>A
OXNAD1	NM_138381	T-j	not performed	synonymous	16312495	
P4HTM	NM_177938	T-g	not performed	missense	49039984	NM_177939.2 (P4HTM): c.679A>G
PBRM1	NM_018165	T-f	not performed	3'UTR	52581406	
PBRM1	NM_018165.4	T-d	Confirmed	Frameshift	52598181	NM_018165.4 (PBRM1): c.3664dup
PBRM1	NM_018165.4	T-e	Confirmed	Frameshift	52643954	NM_018165.4 (PBRM1): c.1846del
PBRM1	NM_018165.4	T-f	Confirmed	Frameshift	52651509	NM_018165.4 (PBRM1): c.1491dup
PBRM1	NM_018165.4	T-g	Confirmed	Frameshift	52613205	NM_018165.4 (PBRM1): c.3302del
PBRM1	NM_018165.4	T-j	Confirmed	Frameshift	52696231	NM_018165.4 (PBRM1): c.446del

				Frequency	Predictions		
protein level	Coverage	Variant	WT	Variant/WT	AGVGD	SIFT	Polyphen
x	516	T	C	26.4/73.6	x	x	x
x	302	A	G	34.1/65.9	x	x	x
x	362	G	C	80.9/19.1	x	x	x
x	689	A	G	62.3/37.7	x	x	x
x	861	A	G	63.6/36.4	x	x	x
	334	C	T	60.5/39.5	x	x	x
	362	G	C	80.9/19.1	x	x	x
	861	A	G	63.6/36.4	x	x	x
p.His261Arg	583	C	T	53.3/46.7	C0	Affect protein function	x
x	334	C	T	60.5/39.5	x	x	x
	176	A	G	67.6/32.4	x	x	x
x	237	T	C	22.8/77.2	x	x	x
x	200	A	G	29.5/70.5	x	x	x
	264	T	C	32.6/67.4	x	x	x
Asn401Ser	229	C	T	32.8/66.8	C0	Tolerated	Benign
	237	G	G	25.3/74.7	x	x	x
x	395	A	G	28.4/71.6	x	x	x
x	303	C	A	64.4/35.6	x	x	x
p.Asn237Ser	435	G	A	29.7/70.3	C0	Tolerated	Benign
x	436	T	G	38.1/61.9	x	x	x
x	449	A	G	32.3/67.7	x	x	x
x	553	T	C	43.8/56.2	x	x	x
x	403	G	C	33.3/66.7	x	x	x
	208	G	A	24.0/76.0	x	x	x
x	319	T	C	28.5/71.5	x	x	x
Gly1052Ser	521	T	C	23.6/76.4	C0	Affect protein function	Probably damaging
x	163	A	G	60.7/39.3	x	x	x
p.Ile227Val	255	G	A	23.5/76.1	C15	Tolerated	Benign
x	471	T	A	39.9/60.1	x	x	x
p.Ile1222AsnfsX5	219	T	-	37.9/61.6	x	x	x
p.Ser616LeufsX8	247	-	A	41.3/58.7	x	x	x
p.Val498CysfsX3	503	A	-	26.0/73.6	x	x	x
p.Asp1101ValfsX26	386	-	T	42.5/57.5	x	x	x
p.Asn149MetfsX25	388	-	T	19.1/80.4	x	x	x

Supplementary Table S1. Continued

Gene name	Refseq	Tumor	Sanger seq	Mutation type	genomic location	cDNA level
PBRM1	NM_018165.4	T-h	Confirmed	missense	52682449	NM_018165.4 (PBRM1): c.724T>C
PBRM1	NM_018165.4	T-a	Confirmed	Splice	52620705	NM_018165.4 (PBRM1): c.3028-1G>A
PDE12	NM_177966	T-b	not performed	missense	57545301	NM_177966.5 (PDE12): c.1400G>A
POU1F1	NM_000306	T-g	not performed	intronic	87325431	
PRICKLE2	NM_198859	T-a	not performed	intronic	64132479	
PRKAR2A	NM_004157	T-d	not performed	missense	48810437	NM_004157.2 (PRKAR2A): c.647C>T
PROK2	NM_001126128	T-h	not performed	3'UTR	71821252	
PROS1	NM_000313	T-h	not performed	3'UTR	93592396	
PROS1	NM_000313	T-e	not performed	missense	93629525	NM_000313.3 (PROS1): c.284G>A
PTPN23	NM_015466	T-a	not performed	3'UTR	47454928	
PTPRG	NM_002841	T-j	not performed	synonymous	62118314	
QRICH1	NM_017730	T-g	not performed	intronic	49084697	
RAF1	NM_002880	T-e	not performed	synonymous	12626130	
RAF1	NM_002880	T-i	not performed	synonymous	12650270	
RBM5	NM_005778	T-c	not performed	intronic	50151010	
RFTN1	NM_015150	T-e	not performed	3'UTR	16357592	
RYBP	NM_012234	T-d	not performed	3'UTR	72424763	
RYBP	NM_012234	T-i	not performed	3'UTR	72424633	
SACM1L	NM_014016	T-c	not performed	synonymous	45751040	
SATB1	NM_001131010	T-b	not performed	intronic	18466151	
SCN10A	NM_006514	T-e	Not confirmed	missense	38770198	NM_006514.2 (SCN10A): c2475A>C
SCN10A	NM_006514	T-f	Not confirmed	intronic	38752461	
SEC13	NM_001136232	T-d	not performed	intronic	10356992	
SEC22C	NM_004206	T-g	not performed	intronic	42597410	
SEMA3F	NM_004186	T-e	not performed	3'UTR	50225794	
SEMA3F	NM_004186	T-e	not performed	intronic	50221027	
SEMA3F	NM_004186	T-a	not performed	missense	64132479	NM_004186.3 (SEMA3F): c.980C>T
SETD2	NM_014159.6	T-g	not performed	synonymous	47162229	
SLC22A14	NM_004803.3	T-h	Not confirmed	Frameshift	38349147	NM_004803.3 (SLC22A14): c711-714del
SMARCC1	NM_003074	T-b	not performed	intronic	47712210	
SNRK	NM_001100594	T-b	not performed	3'UTR	43390102	
SNRK	NM_001100594	T-g	not performed	3'UTR	43391630	
SRGAP3	NM_001033117.1	T-b	not performed	missense	9034594	NM_001033117.1 (SRGAP3): c.2482G>A
SUMF1	NM_182760	T-i	not performed	intronic	4459590	
SUSD5	NM_015551	T-e	not performed	missense	33195202	NM_015551.1 (SUSD5): c.922A>G

				Frequency	Predictions		
protein level	Coverage	Variant	WT	Variant/WT	AGVGD	SIFT	Polyphen
p.Tyr242His	194	G	A	35.1/64.9	C0	Affect protein function	Probably damaging
Splice	381	T	C	13.9/86.1	x	x	x
Arg467His	263	A	G	66.5/33.5	C25	Affect protein function	Probably damaging
	498	G	T	31.5/68.5	x	x	x
	199	A	T	35.2/64.8	x	x	x
Pro216Leu	306	A	G	62.7/37.3	C65	Affect protein function	Probably damaging
x	350	T	C	69.1/30.9	x	x	x
x	397	T	C	50.1/49.9	x	x	x
Gly95Glu	208	T	C	38.0/62.0	C0	Affect protein function	Probably damaging
x	314	T	C	26.8/73.2	x	x	x
x	469	T	C	60.1/39.9	x	x	x
	263	C	T	67.7/32.3	x	x	x
x	164	C	T	75.6/24.4	x	x	x
x	293	C	T	63.1/36.9	x	x	x
	198	T	G	25.3/74.7	x	x	x
x	814	A	G	16.7/84.3	x	x	x
x	434	C	G	22.8/77.2	x	x	x
x	655	A	G	67.3/32.7	x	x	x
x	343	A	G	67.1/32.9	x	x	x
	219	G	C	21.0/78.1	x	x	x
p.Glu825Asp	377	G	T	15.6/84.4	C0	Tolerated	Benign
	165	A	G	69.1/30.9	x	x	x
	226	T	C	23.5/76.5	x	x	x
	274	T	C	64.2/35.4	x	x	x
x	211	A	G	27.5/72.5	x	x	x
	154	T	C	20.1/79.9	x	x	x
p.Pro327Leu	153	T	C	58.2/41.8	C15	Affect protein function	Possibly damaging
x	667	A	G	63.3/36.7	x	x	x
p.Leu238SerfsX25	257	----	TTTG	54.5/44.7	x	x	x
	203	G	A	64.5/35.5	x	x	x
x	351	A	G	15.4/84.6	x	x	x
x	474	T	A	66.7/33.3	x	x	x
Gly828Ser	394	T	C	64.0/36.0	C55	Affect protein function	Possibly damaging
	151	A	G	57.0/43.0	x	x	x
Lys308Glu	356	C	T	31.5/68.5	C0	Affect protein function	Possibly damaging

Supplementary Table S1. Continued

Gene name	Refseq	Tumor	Sanger seq	Mutation type	genomic location	cDNA level
SYN2	NM_133625	T-a	not performed	3'UTR	12232518	
SYN2	NM_003178	T-h	not performed	3'UTR	12226293	
TATDN2	NM_014760	T-h	not performed	3'UTR	10321578	
TGM4	NM_003241	T-c	not performed	3'UTR	44955493	
THRB	NM_000461	T-g	not performed	3'UTR	24164014	
TMEM111	NM_018447	T-f	not performed	synonymous	10016159	
TOP2B	NM_001068	T-g	not performed	intronic	25640770	
TRAIP	NM_005879	T-g	not performed	intronic	49885642	
TRAIP	NM_005879	T-g	not performed	synonymous	49881339	
TRAK1	NM_001042646.1	T-f	not performed	In frame ins	42251578	NM_001042646.1 (TRAK1): c.1963+128_1963+130del
TRANK1	NM_014831.2	T-b	not performed	missense	36871124	NM_014831.2 (TRANK1): c.8608A>G
TTC21A	NM_001105513	T-c	not performed	intronic	39150543	
VGLL3	NM_016206	T-a	not performed	3'UTR	86994861	
VGLL3	NM_016206	T-i	not performed	3'UTR	86994499	
VGLL4	NM_001128220	T-a	not performed	5'UTR	11623819	
VGLL4	NM_001128219	T-a	not performed	intronic	11623819	
VHL	NM_000551.2	T-b	Confirmed	Frameshift	10191498	NM_000551.2 (VHL): c.491_492del
VHL	NM_000551.2	T-e	Confirmed	Frameshift	10188202	NM_000551.2 (VHL): c.346dup
VHL	NM_000551.2	T-g	Confirmed	Frameshift	10188258	NM_000551.2 (VHL): c.401_404del
VHL	NM_000551.2	T-a	Confirmed	missense	10188206	NM_000551.2 (VHL): c.349T>G
VHL	NM_000551.2	T-j	Confirmed	Splice	10188320	NM_000551.2 (VHL): c.463+1del
XCR1	NM_001024644	T-i	not performed	missense	46062889	NM_005283.2 (XCR1): c.551T>C
XYLB	NM_005108	T-d	not performed	missense	38401831	NM_005108.3 (XYLB): c.142A>G
XYLB	NM_005108	T-d	not performed	synonymous	38416748	
ZDHHC3	NM_001135179	T-h	not performed	3'UTR	44962767	
ZDHHC3	NM_001135180	T-h	not performed	intronic	44962767	
ZFYVE20	NM_022340	T-j	not performed	3'UTR	15112960	
ZFYVE20	NM_022340	T-d	not performed	intronic	15126593	
ZNF167	NM_018651	T-a	not performed	3'UTR	44612919	
ZNF167	NM_025169	T-a	not performed	intronic	44612919	
ZNF197	NM_006991.3	T-f	Confirmed	Frameshift	44683961	NM_006991.3 (ZNF197): c.1339_1342del
ZNF385D	NM_024697	T-a	not performed	intronic	21462942	NM_024697.2 (ZNF385D): c.955-3C>A
ZNF445	NM_181489	T-b	not performed	missense	44488293	NM_181489.5 (ZNF445): c.2870A>G

				Frequency	Predictions		
protein level	Coverage	Variant	WT	Variant/WT	AGVGD	SIFT	Polyphen
x	802	T	C	40.4/59.6	x	x	x
x	216	T	C	27.3/72.7	x	x	x
x	262	G	A	54.6/45.4	x	x	x
x	304	A	T	72.0/28.0	x	x	x
x	628	T	C	20.7/79.3	x	x	x
x	236	T	C	28.4/81.6	x	x	x
	164	G	A	38.4/61.1	x	x	x
	442	C	A	31.7/68.3	x	x	x
x	591	T	C	30.6/69.4	x	x	x
225	---	GGA	26.2/73.3	x	x	x	
Ile2870Val	165	C	T	67.3/32.1	C0	Tolerated	Benign
	315	A	T	67.3/32.7	x	x	x
x	330	T	A	20.9/79.1	x	x	x
x	348	A	C	62.6/37.4	x	x	x
x	237	A	C	24.1/75.9	x	x	x
	237	A	C	24.1/75.9	x	x	x
p.Gln164ArgfsX9	167	--	AG	46.7/53.3	x	x	x
p.Leu116ProfsX16	395	C	-	40.0/57.7	x	x	x
p.Glu134AspfsX24	431	----	AATT	42.5/56.8	x	x	x
Trp117Gly	540	G	T	15.2/84.4	C65	Affect protein function	Probably damaging
Splice	253	-	G	22.9/77.1	x	x	x
p.Leu184Pro	577	G	A	68.8/31.2	C0	Tolerated	Probably damaging
p.Thr48Ala	343	G	A	20.4/79.6	C0	Tolerated	Probably damaging
x	373	A	G	27.3/72.7	x	x	x
x	292	T	C	69.9/30.1	x	x	x
	292	T	C	69.9/30.1	x	x	x
x	266	G	T	33.5/66.5	x	x	x
	165	T	G	70.9/29.1	x	x	x
x	545	G	A	63.7/36.3	x	x	x
	545	G	A	63.7/36.3	x	x	x
p.Ile447ThrfsX22	532	----	ATCC	26.3/73.3	x	x	x
	218	T	G	27.5/72.5	x	x	x
Glu957Gly	463	C	T	33.5/66.5	C0	Affect protein function	Benign

Supplementary Table S2. Identified mutations in VHL, PBRM1 and SETD2 in ccRCC-derived cell lines.

cRCC-derived cell line	Gene name	RefSeq ID	Mutation		
			type	cDNA level	protein level
RCC-4	PBRM1	NM_018165.4	Frameshift	c.2592dup	p.His865ThrfsX4
RCC-ER	PBRM1	NM_018165.4	Frameshift	c.753_756del	p.Asp251GlufsX31
RCC-FG2	PBRM1	NM_018165.4	Nonsense	c.1333C>T	p.Gln445X
RCC-HS	PBRM1	NM_018165.4	missense	c.3386G>T	p.Gly1129Val
RCC-MF	PBRM1	NM_018165.4	Frameshift	c.1487del	p.Asn496MetfsX41
RCC-AB	SETD2*	NM_014159.6	Frameshift	c.6447delC	R2132fsX13
RCC-ER	SETD2*	NM_014159.6	Nonsense	c.2017T>A	L655X
RCC-FG2	SETD2*	NM_014159.6	loss of last 3 exons	unknown	unknown
RCC-JF	SETD2*	NM_014159.6	synonymous	c.4189T>C	none
RCC-JW	SETD2*	NM_014159.6	missense	c.3829G>T	G1259V
RCC-MF	SETD2*	NM_014159.6	in frame deletion	c.5330_5338delAACACACAC	1760-1762delNTH
RCC-1	VHL	NM_000551.2	Frameshift	c.458-459ins37	p.Pro154CysfsX2
RCC-4	VHL	NM_000551.2	Missense	c.194C>G	p.Ser65Trp
RCC-ER	VHL	NM_000551.2	Frameshift	c.482del	p.Asp143AlafsX16
RCC-FG2	VHL	NM_000551.2	Frameshift	c.444del	p.Phe148LeufsX11
RCC-JW	VHL	NM_000551.2	Inframe insertion	c.328_444insGCATCA	p.His110delinsArgIleAsn
RCC-MF	VHL	NM_000551.2	Missense	c.256C>T	p.Pro86Ser

* According to the *SIFT* prediction program, the identified variant in *PBRM1* in RCC-HS is likely to affect protein function (score 0.00, median sequence conservation=3.17)
* According to the *Polyphen* prediction program, the identified variant in *PBRM1* in RCC-HS is probably damaging
* According to the *AGVGD* prediction program, the identified variant in *PBRM1* in RCC-HS is less likely to interfere with protein function (score C15).
* According to the SIFT prediction program, the identified variant in *VHL* in RCC-HS is likely to affect protein function (score 0.00, median sequence conservation=3.47)
* According to the Polyphen prediction program, the identified variant in *VHL* in RCC-HS is probably damaging
* According to the AGVGD prediction program, the identified variant in *VHL* in RCC-HS likely to interfere with protein function (score C65).
* According to the SIFT prediction program, the identified variant in *VHL* in RCC-MF is tolerated (score 0.23, median sequence conservation=3.47)
* According to the AGVGD prediction program, the identified variant in *VHL* in RCC-MF less likely to interfere with protein function (score C15).
* The mutations in *SETD2* were identified in an earlier screen (*Duns, Cancer Research 2011*)

THE ENTIRE MIR-200 SEED FAMILY IS STRONGLY DEREGULATED IN CLEAR CELL RENAL CELL CANCER COMPARED TO ITS PROGENITOR CELLS, THE PROXIMAL TUBULAR EPITHELIAL CELLS

6

Gerben Duns[1], Anke van den Berg[2], Inge van Duivenbode[1], Cor Giezen[1], Joost Kluiver[2], Harry van Goor[2], Robert M.W. Hofstra[1], Eva van den Berg[1], Klaas Kok[1]

Departments of Genetics[1] and Pathology & Medical Biology[2], University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands.

In revision

ABSTRACT

Despite numerous studies reporting deregulated microRNA (miRNA) and gene expression patterns in clear cell renal cell carcinoma (ccRCC), no direct comparisons have been made to its presumed normal counterpart; the renal proximal epithelial tubular cells (PTEC). The aim of this study was to determine the miRNA expression profiles of ten clear cell renal cell carcinoma-derived cell lines and short-term cultures of PTEC, and to correlate these with their gene expression, and copy-number profiles. Using microarray-based methods, a significantly altered expression level in ccRCC cell lines was observed for 23 miRNAs and 1630 genes. The set of miRNAs with significantly decreased expression levels include all members of the miR-200 family known to be involved in the epithelial to mesenchymal transition (EMT) process. Expression levels of 13 of the 47 validated target genes for the downregulated miRNAs were increased more than two-fold. Our data reinforce the importance of the EMT process in the development of ccRCC.

INTRODUCTION

Clear cell renal cell carcinoma (ccRCC) is the most prevalent subtype of kidney cancer and accounts for 3% of human malignancies (Jemal et al., 2009). ccRCC is characterized by the frequent loss of 3p (van den Berg et al., 1997) and several genes have been postulated to be involved in its pathogenesis. Based on the current literature, it can be hypothesized that, in addition to protein coding genes, microRNAs (miRNAs) will also be involved in the pathogenesis of ccRCC. Several miRNAs have been shown to exert tumor suppressive or oncogenic functions in a tissue-specific manner (Chen, 2005; Garzon et al., 2006; Kent and Mendell, 2006). Aberrant expression of these miRNAs results in differential translational control of numerous target genes. This can result in the deregulation of cellular processes like proliferation, differentiation, programmed cell death, and morphogenesis, thereby contributing to the process of tumor development (Chen, 2005; Garzon et al., 2006; Kent and Mendell, 2006). Several groups have published up- and downregulated miRNAs in primary ccRCC tumors compared to normal kidney tissue (Gottardo et al., 2007; Nakada et al., 2008; Huang et al., 2009; Petillo et al., 2009; Chow et al., 2010; Juan et al., 2010; Slaby et al., 2010; Weng et al., 2010). However, a disadvantage of examining tumor tissue is the admixture of normal cells and intra-tumoral heterogeneity. The same holds true for the renal tissue used as the normal counterpart, as kidneys are complex organs consisting of many different cell types. This, and the fact that ccRCC is thought to arise specifically from the epithelial cells of the proximal tubuli of the kidney (PTEC), makes the outcome of most studies less reliable (Thoenes et al., 1986). These proximal tubular cells can be isolated from kidney tissue and cultured for a limited number of passages, allowing isolation of RNA for profiling studies from pure cells (Clifford et al., 1998). In the case of ccRCC we postulated that the best strategy would be to compare cancer cell lines with short term cultures of these PTEC cells. We adopted this strategy and analyzed gene and miRNA expression profiles of ccRCC-derived cell lines in comparison to short-term cultures of their precursor cells, i.e. PTECs of the kidney. We further correlated the expression profiles with DNA copy numbers.

MATERIALS AND METHODS

Cell lines

The ccRCC cell lines RCC-AB, RCC-ER, RCC-FG2, RCC-HS, RCCJF, RCC-JW, RCC-MF, and RCC-WK were obtained from Cell Line Services (Eppenheim, Germany). The cells, which tested negative for Mycoplasma, bacteria and fungi, were cultured as recommended. Experiments were performed on cells that had undergone six to eight additional passages. The ccRCC cell lines RCC-1, RCC-4, and RCC-5 were established by Dr. C. D. Gerharz (Institute of Pathology, University Hospital, Düsseldorf, Germany). To ensure their authenticity, they were re-karyotyped. All cRCC cell lines were maintained in RPMI 1640 (Sigma). MN160 and PTEC are primary cultures derived from proximal tubular epithelial cells.

miRNA expression

Cells were grown to 70% confluence. Total RNA fractions were isolated using the Trizol reagent (Invitrogen, Carlsbad, CA, USA). RNA quality was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). miRNA transcript abundance was measured using *Human miRNA microarrays (V2)* (Agilent) following the manufacturer's protocols. The array contained probes for a total of 542 human miRNAs. Slides were scanned using the G25052C DNA microarray scanner (Agilent). Array images were analyzed using Agilent feature extraction software (v10.5.1.1) and GeneSpring GX (Agilent). Data were normalized using a 75% percentile shift (Gibcus et al., 2009). Significantly up- and downregulated miRNAs in the ccRCC derived cell lines compared to the PTEC samples were identified using an unpaired t-test followed by Benjamini-Hochberg correction (false discovery rate FDR: 0.05 and a fold change of at least 2 (FC > 2.0)). Validated targets of individual miRNAs were obtained from MIRECORDS (http://mirecords.biolead.org/).

ArrayCGH

DNA was isolated from ccRCC-derived cell lines using standard protocols. ArrayCGH was carried out using Agilent Human Genome CGH Microarray 244a slides (p/n G4411B) following the manufacturer's protocols. Slides were scanned using the G25052C DNA microarray scanner (Agilent Technologies). Array images were analyzed using Agilent feature extraction software (v10.5.1.1) and NEXUS copy number software (V6.0, Biodiscovery, Inc., El Segundo, CA, USA). In NEXUS, chromosomes with an apparent diploid copy number were used for re-centering. We then used these data to determine the integer copy number of the genomic loci of all miRNA.

Gene expression profiling

RNA from ccRCC-derived cell lines and normal renal samples was isolated using the RNeasy Mini Kit (Qiagen, Venlo, Netherlands) Transcript abundance was measured using Human HT-12_V4_Beadchips (Illumina, San Diego, CA, USA). The Lumi package implemented in R (Du et al., 2008; Lin et al., 2008) was used for variance stabilizing transformation followed by robust spline normalization of the microarray data (GEO Series accession number GSE20491). GeneSpring GX 10.0 (Agilent) was used to calculate fold changes between normal renal samples and ccRCC-derived cell lines. Significantly up- and downregulated genes were identified following the same criteria as given for the miRNA expression arrays, i.e. FC > 2.0, FDR < 0.05. Genes were omitted from further analysis if the average expression level in the ccRCC cell lines and in the PTEC cell lines was below a normalized value of 7, which is close to the background. Pathway-enrichment analyses were done using GeneGo (http://www.genego.com/).

RESULTS

Characterization of the ccRCC cell lines

ArrayCGH analysis revealed multiple chromosomal gains and losses in most of the cell lines (Figure 1, Supplementary Table 1). All cell lines showed a reduced copy number for a large part of the short arm of chromosome 3, with a common region of overlap extending 70 Mb from pter. Other recurrently lost regions included segments on 1p, 4q, 9, 11q and 14q. Eight out of ten cell lines carried a homozygous deletion on 9p21, which varied in size from 50 kb to 600 kb and included the CDKN2A and CDKN2B gene loci.

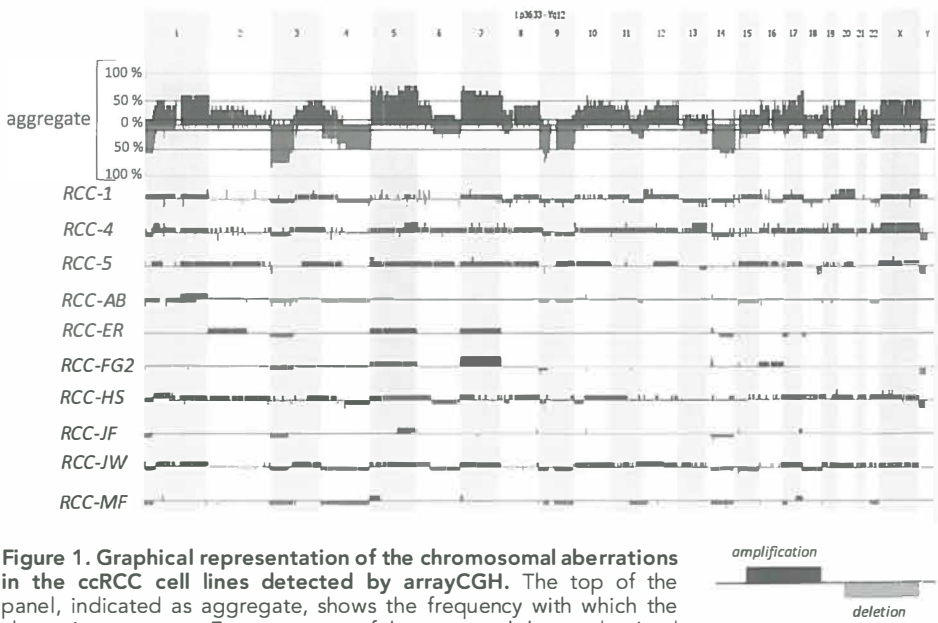


Figure 1. Graphical representation of the chromosomal aberrations in the ccRCC cell lines detected by arrayCGH. The top of the panel, indicated as aggregate, shows the frequency with which the aberrations are seen. For a summary of the recurrently lost and gained regions, see *Supplementary Table 1*.

Gene expression profiling

mRNA profiling revealed a total 11,604 genes (14714 probes) that passed the selection filters. Of these, 767 genes were expressed at significantly higher levels in the ccRCC cell lines compared to PTEC, whereas 869 genes were expressed at significantly lower levels in ccRCC cell lines. This set of 1636 differentially expressed genes was significantly enriched for a number of pathways, several of which were cell cycle-related (supplementary Table 2). Three of the enriched pathways were related to the epithelial-to-mesenchymal transition (EMT). Of the downregulated genes (including the known tumor suppressors VHL and SETD2, but not FHIT) 65 were located at 3p,

the genomic segment most commonly deleted. This equals 22% of all 3p genes that are expressed at a detectable level in PTECs and indicates that genes from 3p are significantly overrepresented in the group of genes that have a decreased abundance in ccRCC (P-value < 0.001).

miRNA expression profiling

Out of 542 human miRNAs that were present on the array, 239 (44%) miRNAs passed the quality filter (flags: present or marginal) in at least four samples. Eight miRNAs showed a significantly increased abundance in ccRCC compared to PTEC and 15 miRNAs showed a significantly decreased abundance (Figure 2a, b, Table 1).

The latter set included all five members of the miR-200 family. Although ccRCC is characterized by recurrent deletions of the short arm of chromosome 3, none of the eight miRNAs that were located at 3p and had detectable levels in PTEC showed a

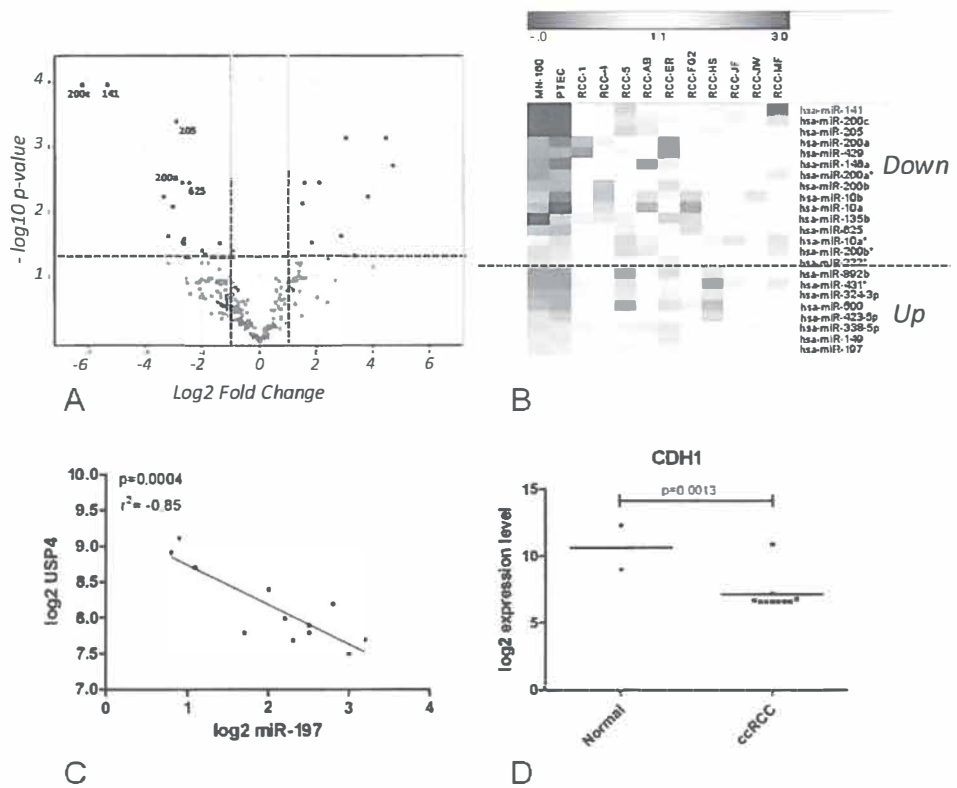


Figure 2. miRNAs showing statistically different levels in ccRCC cell lines compared to normal renal epithelium (FC>2.0, p<0.05). A. Volcano plot, showing average fold changes and corresponding p-values for the individual miRNAs. Dotted lines indicate the thresholds. **B**. Microarray heat map showing the relative abundance of the significantly down- and upregulated miRNAs in the individual cell lines. **C**. Correlation between USP4 transcript levels and miR-197 abundance (Pearson- correlation: -0.85). **D**. Transcript levels of epithelial marker E-cadherin in normal proximal tubular epithelial cells (MN160 and PTEC), and the ccRCC-derived cell lines.

Table 1. miRNAs significantly up- and downregulated in ccRCC samples. Fold change: average expression level in ccRCC-derived cell lines (n=10) versus average expression level in proximal tubular epithelial cells (n=2). Boxed miRNAs are located at genomic regions frequently showing copy number loss.

Upregulated in ccRCC				Downregulated in ccRCC			
miRNA	Locus	Fold change	p-value	miRNA	Locus	Fold change	p-value
miR-892b	Xp27.3	21.8	0.0010	miR-141	12p13.31	73.8	0.0001
miR-431*	14q32.31	14.2	0.0280	miR-200c	12p13.31	40.4	0.0001
miR-324-3p	17p13.1	8.3	0.0010	miR-205	1q32.2	35.3	0.0001
miR-500	Xp11.32	7.3	0.3360	miR-200a	1p36.33	10.2	0.0069
miR-423-5p	17q11.2	4.3	0.0069	miR-429	1p36.33	9.3	0.2800
miR-338-5p	17q25.3	3.6	0.0320	miR-146a	5q33.3	8.3	0.0090
miR-149	2q27.3	3.0	0.0041	miR-200a*	1p36.33	7.5	0.0006
miR-197	1p13.3	2.8	0.0083	miR-200b	1p36.33	6.7	0.0041
				miR-10b	2q31.3	6.5	0.0320
				miR-10a	17q21.31	6.3	0.0337
				miR-135b	1q32.1	6.2	0.0305
				miR-625	14q32.3	5.6	0.0041
				miR-10a*	17q21.31	4.0	0.0426
				miR-200b*	1p36.33	3.7	0.0480
				miR-222*	Xp11.3	2.7	0.0333

Fold change: average expression level in ccRCC-derived cell lines (n=10) vs average expression level in proximal tubular epithelial cells (n=2). Boxed miRNAs are located at genomic regions frequently showing copy number loss.

significant decrease in abundance across the entire panel of ccRCC cell lines. Let-7g and miR-138, both reported to be suppressed in some cancer types, exhibited a more than two-fold decrease in abundance in 6/10 and 8/10 ccRCC cell lines, respectively (Supplementary Table 3). One pathway, development of microRNAs-dependent inhibition of EMT (p-value: 4.6E-19), appeared to be significantly enriched for the set of 23 microRNAs that showed consistent aberrant expression in ccRCC.

Coupling of miRNA expression to copy number aberrations and gene expression levels

An altered level of miRNA expression in the ccRCC cell lines compared to the PTEC cells can be a consequence of an altered copy number for the miRNA locus. To address this question we analyzed if there was a correlation between the integer copy number and levels of expression of the miRNAs. For 73 miRNA loci, copy number aberrations were detected in at least two cell lines. The correlation was > 0.5 for seven miRNAs, while the highest correlation was detected for miR-574 located at 4p14, with a correlation coefficient of 0.74.

To determine the pathogenic consequences of the significantly altered miRNA levels, we correlated their expression levels to those of the protein coding genes. We focused on the proven target genes of the 23 miRNAs we had identified as being significantly deregulated. In total 64 validated targets were reported in the miRNA database, of which 47 passed the quality filters set for the microarray analysis. Interestingly, this set included only one gene, i.e. FUS1/TUSC2, for the set of upregulated miRNAs. The other 46 genes were targets of the downregulated miRNAs. Thirteen of these genes had a significantly higher mRNA level in the ccRCC cell lines, with an absolute fold change > 4 (Supplementary Table 4), supporting their interaction with the respective miRNAs.

miRNAs can exert oncogenic functions by targeting tumor suppressor genes. We focused on 3p, a region of the genome that is deleted in most, if not all, ccRCC (van den Berg et al., 1997). The expression profiling data revealed 65 significantly downregulated transcripts that were located on 3p. Eight of these are predicted by TargetScan to contain a conserved target sequence for one of the significantly upregulated miRNAs (Supplementary Table 5). Of these predicted miRNA/mRNA combinations, miR-197 and ubiquitin specific peptidase 4 *(USP4)* showed the most significant inversed correlation (P-value = -0.85) (Figure 2c).

DISCUSSION

The kidney is a complex organ comprising many different cell types. Based on the variety of functions, each cell type in the kidney will have its own specific miRNA profile. Thus the miRNA profile of a kidney specimen is, to a large part, a reflection of the composition of that particular sample. ccRCC is considered to be derived from a specific type of cell, the epithelial cells of the proximal tubulus. Tumor specimens will also contain variable amounts of different normal cell types, which will all contribute to the miRNA profile of that specimen. As an example, we used an exome sequencing approach on ten primary ccRCCs, all microscopically selected to have at least 80% tumor cells. The resulting data indicated that the fraction of tumor cells was often much lower, in some cases even less than 50%, due to unrecognized infiltrating lymphocytes. In the current study we circumvented these disadvantages by analyzing ccRCC-derived cell lines and comparing their miRNA profiles with short-term cultures of epithelial cells of the proximal tubulus of the kidney.

Fifteen miRNAs were significantly downregulated in all ten ccRCC cell lines compared to PTEC. We are the first to show such decreased levels for all the miRNA-200 family members and miR-205. In reported studies on primary ccRCC tumor specimens in comparison to the total renal tissue, only individual members of this miRNA family were identified (Nakada et al., 2008; Chow et al., 2010; Slaby et al., 2010; Weng et al., 2010), indicating the strength of our approach using ccRCC-derived cell lines in combination with short-term cultures of PTEC.

A decreased miR-205 level has not been reported before in ccRCC tumors, but has been observed in several other tumor types (Volinia et al., 2006). miR-205 shares some important target genes with the miR-200 family that suppress the epithelial- to

mesenchymal transition (EMT) process (Gregory et al., 2008; review by Mongroo and Rustgi , 2010). miR-200 family members regulate the processes of EMT mainly by regulating ZEB1 and ZEB2 translation through binding to their 3'UTR sequences. ZEB1 and ZEB2 are part of a complex network of transcriptional repressors that regulate the expression of E-cadherin and a number of master regulators of epithelial polarity. Thus, a decreased abundance of the miR-200 family will result in an increased activity of *ZEB1* and *ZEB2*, amongst others resulting in a repression of E-cadherin. Indeed, the transcript level of E-cadherin is strongly decreased in 9/10 of our ccRCC-derived cell lines (Figure 2d). Three EMT-related pathways were significantly enriched for the differentially expressed genes (Supplementary Table 2), further supporting the importance of this transition in the development of ccRCC. The development of the proximal tubuli of the kidney is the product of a complex series of events during embryogenesis that include a transition from mesenchyme to epithelium (Davies et al., 1999). The development of ccRCC from these progenitor cells thus appears to include the reciprocal process. The acquisition of a mesenchymal phenotype gives tumor cells the capacity to infiltrate surrounding tissues, leading to increased cell motility and invasion. As such, the process appears to be linked to tumor progression rather than to initiating steps.

Among the other significantly decreased miRNAs are miR-146, loss of which results in leukemia in a mouse model (Starczynowski et al., 2011), and miR-10a and miR-10b, which are involved in the transcriptional regulation of HOX genes and are deregulated in several cancer types (Lund, 2010). In studies on primary tumor specimens, miR-200a* and the primary transcripts of miR-10a and miR-10b correlated with a good prognosis (Petillo et al., 2009). miR-10a and 10b have also been associated with non-metastatic ccRCC (Heinzelmann et al., 2011). miR-324, miR-423-5p and miR-149, which showed significantly decreased levels in our ccRCC-derived cell lines, have been reported to discriminate between ccRCC and papillary RCC (Petillo et al., 2009). Three other miRNAs, i.e. miR-200b, miR-200c, and miR10b are part of a 65-member miRNA-signature that discriminates between different types of kidney tumors (Youssef et al., 2011).

None of the eight miRNAs that showed significantly increased levels in the ccRCC-derived cell lines compared to PTEC have been reported in ccRCC before. Only miR-197 has a proven target: tumor suppressor candidate 2 (TUSC2/FUS1), located at the 3p21.3 region that is frequently lost (Du et al., 2008). The TUSC2 transcript levels seem to be unaltered in ccRCC. However, this interaction may well be important in the pathogenesis of ccRCC, as, in general, miRNA targeting leads to inhibition of translation and not degradation of its target genes.

According to Knudson's classical model, recurrent loss of specific chromosomal segments in tumors suggests the presence of tumor-suppressive genes in these regions. Such a model could well apply to miRNAs. The clustering of miRNAs in regions of the chromosome that are frequently unstable in cancer indeed suggests that genomic alterations might be one of the mechanisms that control miRNA expression in cancer (Calin and Croce, 2006). Of the miRNAs that showed a decreased transcript abundance in the ccRCC-derived cell lines, six were located in genomic loci that are

frequently lost in ccRCC, i.e. miR-200a, miR-200a*, miR-200b, miR-200b* and miR-429 located at 1p36.33, and miR-625 located at 14q23.3. This suggests that the loss of the chromosomal region that contains a miRNA locus is an important factor in the functional loss of these miRNAs.

None of the 15 miRNAs that were consistently downregulated map to the short arm of chromosome 3, although let-7g and miR-138 did have a decreased level in 6/10 and 8/10 of the analyzed cell lines, respectively. Thus, loss of function of one of the known miRNAs from 3p does not appear to be the driving force for the frequently observed loss of this chromosomal segment. As reduced expression of miRNAs by chromosomal losses explains only a subset of cases, other mechanisms such as epigenetic changes, or mutations within the miRNA, may play a role in the inactivation of these miRNAs.

In conclusion, we have used an alternative approach to identify miRNAs whose expression is consistently altered in ccRCC compared to their normal counterpart, the epithelial cells of the proximal tubulus, and pinpointed the epithelial- to mesenchymal transition as an important event in the development of ccRCC. The data presented here will help to interpret and understand the findings in primary RCC specimens, and may help to elucidate some of the processes that lead to the development and progression of ccRCC. With the availability of detailed expression-, miRNA- and CGH profiles, these cell lines should prove valuable in further studies.

ACKNOWLEDGEMENTS

This work was supported by Grant RUG-2007-3892 from the Dutch Cancer Society and by a Grant from the Jan Kornelis De Cock foundation in Groningen. We thank Jackie Senior for editing the manuscript.

REFERENCES

Calin GA, Croce CM. 2006. MicroRNAs and chromosomal abnormalities in cancer cells. Oncogene 25:6202-6210.

Chen CZ. 2005. MicroRNAs as oncogenes and tumor suppressors. N Engl J Med 353:1757-1771.

Chow TF, Mankaruos M, Scorilas A, Youssef Y, Girgis A, Mossad S, Metias S, Rofael Y, Honey RJ, Stewart R, Pace KT, Yousef GM. 2010. The miR-17-92 Cluster is Over Expressed in and Has an Oncogenic Effect on Renal Cell Carcinoma. J Urol 183:743-751.

Clifford SC, Czapla K, Richards FM, O'Donoghue DJ, and Maher ER. 1998. Hepatocyte growth factor-stimulated renal tubular mitogenesis: effects on expression of c-myc, c-fos, c-met, VEGF and the VHL tumour-suppressor and related genes. Br J Cancer 77:1420–1428.

Davies JA, Perera AD and Walker CL. 1999. Mechanisms of epithelial development and neoplasia in the metanephric kidney. Int J Dev Biol 43: 473-478.

Du P, Kibbe WA, Lin SM. 2008. Lumi: a pipeline for processing Illumina microarray. Bioinformatics 59:1547-1548.

Garzon R, Fabbri M, Cimmino A, Calin GA, Croce CM. 2006. MicroRNA expression and function in cancer. Trends Mol Med 12:580-587.

Gibcus JH, Tan LP, Harms G, Schakel RN, de Jong D, Blokzijl T, Möller P, Poppema S, Kroesen BJ, van den Berg A. 2009. Hodgkin lymphoma cell lines are characterized by a specific miRNA expression profile. Neoplasia 11:167-176.

Gottardo F, Liu CG, Ferracin M, Calin GA, Fassan M, Bassi P, Sevignani C, Byrne D,

Negrini M, Pagano F, Gomella LG, Croce CM, Baffa R. 2007. Micro-RNA profiling in kidney and bladder cancers. Urologic oncology 25:287-292.

Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nature Cell Biology 10:593-601.

Heinzelmann J, Henning B, Sanjmyatav J, Posorki N, Steiner T, Wunderlich H, Gadja MR, Junker K. 2011. Specific miRNA signatures are associated with metastasis and poor prognosis in clear cell renal cell carcinoma. World J Urol 2011;29:367-373.

Huang Y, Dai Y, Yang J, Chen T, Yin Y, Tang M, Hu C, Zhang L. 2009. Microarray analysis of microRNA expression in renal clear cell carcinoma. Eur J Surg Oncol 35:1119-1123.

Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. 2009. Cancer statistics, 2009. CA Cancer J Clin 59:225-249.

Kent OA, Mendell JT. 2006. A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. *Oncogene 25:6188-6196.*

Juan D, Alexe G, Antes T, Liu H, Madabhushi A, Delisi C, Ganesan S, Bhanot G, Liou LS. 2010. Identification of a microRNA panel for clear-cell kidney cancer. Urology 75:835-841.

Lin SM, Du P, Huber W, Kibbe WA. 2008. Model-based variance- stabilizin transformation for Illumina microarray data. Nucleic Acids Res 36:e11.

Lund AH. 2010. miR-10 in development and cancer. Cell Death Differ 17:209-214.

Mongroo PS and Rustgi AK. 2010. The role of the miR-200 family in epithelial-mesenchymal transition. Cancer Biology & Therapy 10: 219-222.

Nakada C, Matsuura K, Tsukamoto Y, Tanigawa M, Yoshimoto T, Narimatsu T, Nguyen LT, Hijiya N, Uchida T, Sato F, Mimata H, Seto M, Moriyama M. 2008. Genome-wide microRNA expression profiling in renal cell carcinoma: significant down-regulation of miR-141 and miR-200c. J Pathol 216:418-427.

Petillo D, Kort EJ, Anema J, Furge KA, Yang XJ, Teh BT. 2009. MicroRNA profiling of human kidney cancer subtypes. Int J Oncol;35:109-114.

Slaby O, Jancovicova J, Lakomy R, Svoboda M, Poprach A, Fabian P, Kren L, Michalek J, Vyzula R. 2010. Expression of miRNA-106b in conventional renal cell carcinoma is a potential marker for prediction of early metastasis after nephrectomy. J Exp Clin Cancer Res 29:90.

Starczynowski DT, Morin R, McPherson A, Lam J, Chari R, Wegrzyn J, Kuchenbauer F, Hirst M, Tohyama K, Humphries RK, Lam WL, Marra M, Karsan A. 2011. Genome-wide identification of human microRNAs located in leukemia-associated genomic alterations. Blood 117:595-607.

Thoenes W, Störkel S, Rumpelt HJ. 1986. Histopathology and classification of renal cell tumors (adenomas, oncocytomas and carcinomas). The basic cytological and histopathological elements and their use for diagnostics. Pathol Res Pract 181:125-143.

Van den Berg A, Dijkhuizen T, Draaijers TG, Hulsbeek MM, Maher ER, Van den Berg E, Störkel S, Buys CH. 1997. Analysis of multiple renal cell adenomas and carcinomas suggests allelic loss at 3p21 to be a prerequisite for malignant development. Genes Chromosomes Cancer 19:59–76.

Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM. 2006. A miRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci USA 103:2257-2261.

Weng L, Wu X, Gao H, Mu B, Li X, Wang JH, Guo C, Jin JM, Chen Z, Covarrubias M, Yuan YC, Weiss LM, Wu H. 2010. MicroRNA profiling of clear cell renal cell carcinoma by whole-genome small RNA deep sequencing of paired frozen and formalin-fixed, paraffin-embedded tissue specimens. J Pathol 222:41-51.

Youssef YM, White NM, Grigull J, Krizova A, Samy C, Mejia-Guerrero S, Evans A, Yousef GM. 2011. Accurate molecular classification of kidney cancer subtypes using microRNA signature. Eur Urol 59:721-730.

SUPPLEMENTARY FIGURES AND TABLES

Supplementary Table 1. Recurrent copy number alterations in the ccRCC-derived cell lines

	Copy number losses			Copy number gains	
Location	Region of overlap *(Mb from pter)*	Frequency	Location	Region of overlap *(Mb from pter)*	Frequency
1p	pter-28,099	7/10	5q	129,266-qter	6/10
3p	pter-64,696	10/10	7p (1)	3,934-11,673	5/10
4q	84,724-qter	5/10	7p (2)	27,174-cen	6/10
9p	22,022-31,998	8/10	7q	cen-qter	5/10
9q	cen-qter	5/10			
11q	112,852-qter	5/10			
14q	49,977-qter	7/10			

Supplementary Table 2. Pathways significantly represented in the group of up- and down-regulated genes.

Rank	Pathway Maps	pValue	Ratio
1	Cell cycle_Chromosome condensation in prometaphase	1.057E-13	15/21
2	Cell cycle_The metaphase checkpoint	1.609E-13	19/36
3	Cell cycle_Role of APC in cell cycle regulation	2.734E-12	17/32
4	Cell cycle_Initiation of mitosis	4.919E-12	15/25
5	Cell cycle_Spindle assembly and chromosome separation	7.444E-11	16/33
6	DNA damage_ATM/ATR regulation of G1/S checkpoint	5.419E-10	15/32
7	Cell adhesion_Chemokines and adhesion	1.730E-09	26/100
8	Cytoskeleton remodeling_Cytoskeleton remodeling	1.323E-08	25/102
9	DNA damage_ATM / ATR regulation of G2 / M checkpoint	3.769E-08	12/26
10	Cell cycle_Start of DNA replication in early S phase	6.580E-08	13/32
11	Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling	8.085E-08	25/111
12	Cell cycle_Transition and termination of DNA replication	1.045E-07	12/28
13	Cell cycle_Sister chromatid cohesion	6.392E-07	10/22
14	Cell cycle_Regulation of G1/S transition (part 1)	6.991E-07	13/38
15	Cell adhesion_ECM remodeling	1.185E-06	15/52
16	Reproduction_Progesterone-mediated oocyte maturation	1.368E-06	13/40
17	Immune response_Histamine H1 receptor signaling in immune response	2.329E-06	14/48
18	Cell cycle_Cell cycle (generic schema)	4.337E-06	9/21
19	Cell cycle_Role of Nek in cell cycle regulation	4.836E-06	11/32
20	Cell cycle_ESR1 regulation of G1/S transition	6.811E-06	11/33

Supplementary Table 2. continued

Rank	Pathway Maps	pValue	Ratio
21	dCTP/dUTP metabolism	8.903E-06	17/75
22	Apoptosis and survival_p53-dependent apoptosis	1.276E-05	10/29
23	Cell cycle_Nucleocytoplasmic transport of CDK/Cyclins	1.529E-05	7/14
24	Signal transduction_AKT signaling	2.034E-05	12/43
25	Immune response_Oncostatin M signaling via MAPK in human cells	2.339E-05	11/37
26	Signal transduction_cAMP signaling	3.091E-05	11/38
27	Cell adhesion_Endothelial cell contacts by junctional mechanisms	3.379E-05	9/26
28	Cell cycle_Role of 14-3-3 proteins in cell cycle regulation	6.230E-05	8/22
29	Immune response_MIF-mediated glucocorticoid regulation	6.230E-05	8/22
30	Cell adhesion_Integrin-mediated cell adhesion and migration	6.733E-05	12/48
31	Immune response_Oncostatin M signaling via MAPK in mouse cells	8.040E-05	10/35
32	Cell cycle_Role of SCF complex in cell cycle regulation	8.995E-05	9/29
33	Cell adhesion_Tight junctions	1.046E-04	10/36
34	Apoptosis and survival_Granzyme A signaling	1.208E-04	9/30
35	DNA damage_Brca1 as a transcription regulator	1.208E-04	9/30
36	DNA damage_Role of Brca1 and Brca2 in DNA repair	1.208E-04	9/30
37	DNA damage_DNA-damage-induced responses	1.479E-04	5/9
38	Immune response_IL-17 signaling pathways	1.647E-04	13/60
39	Development_WNT signaling pathway. Part 2	1.881E-04	12/53
40	Development_TGF-beta-dependent induction of EMT via RhoA, PI3K and ILK.	2.073E-04	11/46
41	Cell adhesion_Cadherin-mediated cell adhesion	2.383E-04	8/26
42	Development_TGF-beta-dependent induction of EMT via MAPK	2.543E-04	11/47
43	Immune response_CD40 signaling	3.251E-04	13/64
44	Development_Regulation of epithelial-to-mesenchymal transition (EMT)	3.251E-04	13/64
45	Apoptosis and survival_HTR1A signaling	4.520E-04	11/50
46	Regulation of CFTR activity (norm and CF)	4.590E-04	12/58
47	Development_Beta-adrenergic receptors transactivation of EGFR	6.821E-04	9/37
48	ATP/ITP metabolism	7.224E-04	19/124
49	Regulation of lipid metabolism_Regulation of lipid metabolism by niacin and isoprenaline	7.501E-04	10/45
50	Cell adhesion_Histamine H1 receptor signaling in the interruption of cell barrier integrity	7.501E-04	10/45

Ratio: number of genes with a significantly altered level of expression versus the total number of genes in this pathway as indicated in the GeneGo package

Supplementary Table 3. miRNAs located in recurrently lost regions that show an average decrease in transcript abundance of two-fold or more in a subset of the ccRCC-derived cell lines.

miRNA	Locus	RCC-1	RCC-4	RCC-5	RCC-AB
hsa-miR-200b	1p36.33	7.6 ↓	27.0 ↓	4.0 ↓	4.3 ↓
hsa-miR-200b*	1p36.33	2.6 ↓	4.8 ↓	1.1 ↑	9.2 ↓
hsa-miR-200a*	1p36.33	5.0 ↓	4.9 ↓	10.7 ↓	7.4 ↓
hsa-miR-200a	1p36.33	40.3 ↓	16.4 ↓	4.1 ↓	5.8 ↓
hsa-miR-429	1p36.33	78.2 ↓	8.3 ↓	2.8 ↓	5.2 ↓
hsa-miR-34a	1p36.22	2.7 ↓	4.5 ↓	1.1 ↓	3.2 ↓
hsa-let-7g	3p21.1	5.4 ↓	45.3 ↓	1.7 ↓	2.5 ↓
hsa-miR-138	3p21.32 & 16q13	18.5 ↓	3.2 ↓	4.7 ↓	5.9 ↓
hsa-let-7f	9q22.32 & Xp11.22	6.0 ↓	60.1 ↓	3.1 ↓	1.6 ↓
hsa-let-7d	9q22.32	3.0 ↓	23.7 ↓	1.8 ↓	1.4 ↓
hsa-miR-23b	9q22.32	4.6 ↓	10.2 ↓	1.3 ↓	1.3 ↑
hsa-miR-27b	9q22.32	8.7 ↓	9.3 ↓	1.6 ↓	1.0 –
hsa-miR-21	9q31.3	6.0 ↓	11.3 ↓	1.2 ↓	1.5 ↓
hsa-let-7a	9q22.32 & 11q24.1 & 22q13.31	4.8 ↓	33.7 ↓	3.7 ↓	1.4 ↓
hsa-miR-625	14q23.3	3.0 ↓	7.9 ↓	11.9 ↓	8.1 ↓
hsa-miR-134	14q23.3	4.6 ↓	3.4 ↓	2.1 ↓	2.3 ↓

Numbers indicate the absolute fold change for the ccRCC-derived cell line relative to PTEC. (↑) positive FC, (↓) negative FC. Total# indicates number of cell lines with a fold change > 2.

Supplementary Table 4. Validated target genes with a significantly higher mRNA level in the ccRCC cell lines.

Gene symbol	Fold Change	microRNA
BLMH	3.2	miR-146a
BRCA1	2.1	miR-146a
CCNA2	4.7	miR-146a
CDKN3	3.4	miR-146a
HMGB1	2.9	miR-141
MCM10	5.7	miR-146a
NFIX	2.5	miR-146a
PA2G4	3.2	miR-146a
POLE2	2.4	miR-146a
RAD54L	2.9	miR-146a
TIMELESS	2.5	miR-146a
UHRF1	5.0	miR-146a
USF2	2.1	miR-10a

RCC-ER	RCC-FG2	RCC-HS	RCC-JF	RCC-JW	RCC-MF	Total #
18.7 ↓	3.1 ↓	7.4 ↓	4.2 ↓	6.6 ↓	6.9 ↓	10
1.8 ↓	3.6 ↓	5.8 ↓	4.6 ↓	2.7 ↓	9.0 ↓	9
3.5 ↓	6.8 ↓	11.3 ↓	9.6 ↓	5.3 ↓	17.5 ↓	10
74.8 ↓	4.4 ↓	8.5 ↓	7.1 ↓	9.2 ↓	6.4 ↓	10
68.5 ↓	4.0 ↓	4.5 ↓	6.7 ↓	10.5 ↓	7.6 ↓	10
19.4 ↓	2.2 ↓	2.7 ↓	1.8 ↓	2.9 ↓	1.6 ↓	7
12.5 ↓	1.6 ↓	16.8 ↓	1.5 ↓	16.4 ↓	1.5 ↓	6
13.1 ↓	7.6 ↓	2.0 ↑	2.8 ↓	1.0 –	10.5 ↓	8
6.0 ↓	1.0 –	23.6 ↓	1.1 ↑	28.4 ↓	1.3 ↓	6
4.0 ↓	1.0 –	11.3 ↓	1.1 ↑	13.1 ↓	1.1 –	5
4.5 ↓	1.0 –	2.6 ↓	1.1 ↑	21.0 ↓	1.3 ↓	5
6.1 ↓	1.2 ↓	2.8 ↓	1.3 ↓	13.9 ↓	1.5 ↓	5
4.2 ↓	1.6 ↓	13.7 ↓	2.4 ↓	6.5 ↓	1.1 ↓	6
3.6 ↓	1.2 ↑	17.0 ↓	1.2 ↑	34.7 ↓	1.5 ↓	6
3.0 ↓	1.5 ↓	9.6 ↓	8.2 ↓	4.5 ↓	6.0 ↓	9
4.0 ↓	7.6 ↓	2.3 ↑	1.7 ↓	3.2 ↓	4.0 ↓	9

Supplementary Table 5. Predicted targets on 3p for significantly upregulated miRNAs.

microRNA				predicted target			Pearson correlation
name	locus	FC	p-value	Symbol	FC	p-value	
hsa-miR-149	2q27.3	3.0 ↑	0.0041	ARPC4	4.9 ↓	6.51E-09	-0.63
				OXSR1	2.9 ↓	6.60E-09	-0.57
hsa-miR-197	1p13.3	2.8 ↑	0.0083	USP4	2.2 ↓	2.25E-07	-0.85
hsa-miR-324-3p	17p13.1	8.3 ↑	0.0010	GNAI2	3.0 ↓	1.15E-09	-0.70
				SUMF1	2.6 ↓	9.13E-11	-0.81
hsa-miR-338-5p	17q25.3	3.6 ↑	0.0320	VGLL4	2.6 ↓	9.37E-06	-0.75
				GNAI2	3.0 ↓	1.15E-09	-0.57
hsa-miR-892b	Xp27.3	21.8 ↑	0.0010	UBP1	2.0 ↓	2.97E-09	-0.72

FC: Fold change (ccRCC versus PTEC). (↑) positive FC, (↓) negative FC

7

GENERAL DISCUSSION

GENERAL DISCUSSION

In this thesis, we describe different approaches used to identify novel ccRCC tumor suppressor genes. Here, I summarize our findings, put them in a broader perspective, and speculate on future perspectives.

Chromosomal aberrations and *VHL* mutations in ccRCC

To facilitate our search for ccRCC tumor suppressor genes, we obtained a panel of ccRCC-derived cell lines. During the course of this project, these cell lines were characterized in detail with respect to their chromosomal copy number, miRNA expression, and gene expression profiles. Moreover, these cell lines were screened for inactivating mutations in the *Von Hippel Lindau* (*VHL*) gene, which is a known ccRCC tumor suppressor gene.

We assessed chromosomal copy number aberrations for a total of eleven ccRCC-derived cell lines, and 22 primary ccRCC tumors. We observed, amongst various other aberrations, a complete loss of 3p in seven cell lines and 14 primary tumors. Three cell lines and five tumors had a large terminal deletion, the smallest being 64 Mb. One primary tumor showed retainment of the terminal part (pter to 24 Mb from pter), which included the 3p25 region. The two tumors that did not show 3p loss did not show a loss or gain for any other locus, indicating that these samples may not have contained sufficient tumor cells to detect copy number aberrations. Our findings underscore those of recent papers that show that the vast majority of ccRCC tumors have large terminal deletions of 3p that comprise almost the entire p-arm (Toma et al.,2008; Sükösd et al., 2003). In contrast to what others have observed (Van den Berg et al.,1997; Chudek et al., 1997; Braga et al., 2002), none of the ccRCC samples that we analyzed showed interstitial 3p deletions. One cell line that did not show 3p loss was excluded from subsequent experiments, as this cell line may not hold clues for a 3p-located tumor suppressor gene. Other recurrently lost regions include segments on 1p, 4q, 9, 11q, and 14q. Eight out of ten cell lines carried a homozygous deletion on 9p21 varying in size from 50 to 600 kb, which included the CDKN2A and CDKN2B gene loci, encoding the P16 and P15 proteins, respectively. This suggests a possible role for p15 and/or p16 inactivation in ccRCC development.

We identified inactivating mutations in *VHL*, located at 3p25, in five out of ten ccRCC-derived cell lines and in six out of ten primary ccRCC tumors that were screened for *VHL* mutations. This is in agreement with previous reports of the inactivation of *VHL* in about 50 % of ccRCC tumor samples (Gnarra et al.,1994; Nickerson et al.,2008). The function of VHL and its role in ccRCC tumorigenesis have been well studied during the past decade. VHL is part of a ubiquitin ligase complex that targets transcription factor HIFα for proteosomal degradation. Inactivation of VHL results in consistently high HIFα levels, triggering the upregulation of HIFα targets, which include *vascular endothelial growth factor* (*VEGF*) and *platelet derived growth factor* (*PDGF*). These targets in turn activate pathways such as the *RAF-MEK-ERK* and *Phosphatidylinositol-3 kinase-AKT-mTOR* pathways, triggering angiogenesis and proliferation (Clark et al., 2009; Baldewijn et al., 2010; reviewed by Kaelin, 2008). Recent reports show that *VHL*

is involved in several other processes, including tight junction formation, suggesting a role in epithelial-to-mesenchymal transition (Harten et al., 2009). Although VHL is a major player in ccRCC tumorigenesis, VHL is not inactivated in all ccRCC tumors, and VHL loss alone is not sufficient for ccRCC development, suggesting the need for additional genetic events (Mandriota et al., 2002; Young et al., 2008). For a long time the search for (ccRCC) tumor suppressor genes has been focused on protein coding genes. Recent findings, however, indicate that noncoding genes, including miRNAs, can exert tumor suppressive or oncogenic functions in a tissue-specific manner (reviews by Kent et al., 2006; Garzon et al., 2006; Chen et al., 2005).

miRNAs and ccRCC

In ***Chapter 6***, we compared the genome-wide microRNA expression profiles in the ccRCC-derived samples with those in proximal tubular epithelial cells (PTECs), which are the cells from which ccRCC tumors are thought to be derived (Thoenes et al., 1985). We showed a significant downregulation of all members of the miR-200 family and miR-205. These miRNAs control cellular morphogenesis by translational regulation of targets that include the transcription factors ZEB1 and ZEB2, which in their turn regulate the expression of E-cadherin and a number of master regulators of epithelial polarity (Gregory et al., 2008). As a result of loss of the miR-200 family, epithelial cells undergo an epithelial-to-mesenchymal transition (EMT), which, in a tumorigenic context, confers a metastatic potential to the cells (Gibbons et al., 2009). Thus, our data suggest a role for miRNAs in ccRCC progression. None of the known miRNAs that are located on the short arm of 3p were consistently downregulated in the ccRCC samples, suggesting that loss-of-function of these miRNAs is not the driving force for the frequently observed loss in this chromosomal segment.

A new gene for ccRCC, *SETD2*

In ***Chapter 2***, we applied *Gene Identification of nonsense-mediated mRNA-decay Inhibition (GINI)* on the ccRCC-derived cell lines that showed 3p loss, and identified *SETD2* as a new candidate ccRCC tumor suppressor gene. *SETD2* is located at 3p21, and encodes a histone methyltransferase that is nonredundantly responsible for trimethylation of lysine 36 of histone 3 (H3K36me3). We identified inactivating mutations in five out of ten ccRCC-derived cell lines. These five cell lines, and two other cell lines, showed loss or significant decrease of global H3K36me3 modification levels. Sequencing of *SETD2* in a total of 20 primary ccRCC tumors revealed missense mutations in two of them. A huge exome-sequencing project revealed inactivating mutations in *SETD2* in 12 out of 407 (3%) primary ccRCC tumors, providing *in vivo* evidence for the relevance of our findings (Dalgliesh et al., 2010).

Dalgliesh et al. (2010) did not identify mutations in *SETD2* in other tumor types, and suggested that inactivation of *SETD2* might be specific for ccRCC tumors. We performed a mutational screening of *SETD2* in nine small-cell lung carcinoma-derived cell lines, a type of carcinoma also characterized by loss of 3p21, but also did not find any mutations. However, the number of investigated non-ccRCC tumors might not have been sufficient to give a good estimate of the mutation rate in other tumor

types. Moreover, decreased *SETD2* transcript levels have been observed in malignant breast cancer (Al Sarakbi et al., 2009), and a transposon-based genetic screen in mice suggested a role for *SETD2* inactivation in the development of pancreatic cancer (Starr et al., 2009). Dalgliesh et al. (2010) reported the identification of inactivating mutations in other histone-modifying enzymes in a subset of ccRCC tumors; this suggests a broader role for aberrant histone 3 methylation in ccRCC development. Therefore, inactivating mutations in these different histone H3-modifying enzymes might contribute to tumorigenic development by affecting the same pathways. Aberrant activity of histone- modifying enzymes could result in altered chromatin configuration and disruption of normal transcriptional programs, pushing the cell towards cancerous development (Morin et al., 2011; reviewed by Füllgrabe et al., 2011).

The function of SETD2

The mechanisms through which inactivation of *SETD2* contributes to ccRCC development are not yet fully known. However, one could speculate that loss of H3K36me3, which is the result of *SETD2* inactivation, results in deregulated transcription and/or transcription-related processing of a set of genes that play a role in ccRCC tumorigenesis. Trimethylation of H3K36me3 is intimately linked with transcriptional elongation, as SETD2 is recruited to nucleosomes by hyperphosphorylated RNA polymerase during transcriptional elongation (Yoh et al., 2008). Furthermore, the preferential location of H3K36me3 on exons relative to intronic regions suggests a role for HK36me3 in exon definition and, therefore, in regulating splicing (Kolasinska-Zwierz et al., 2009; Spies et al., 2009, Andersson et al., 2009, Huff et al., 2010). Histone modifications have previously been suggested to be able to modulate the outcome of alternative splicing by influencing the elongation rate of RNAPOLII (Schor et al., 2009). Slowing of the elongation rate could result in recognition of "weak" splice sites, resulting in inclusion of alternative exons (De la Mata et al., 2003). Indeed, recently, Luco et al. (2010) showed that H3K36me3 can directly affect splicing by recruiting polypyrimidine tract-binding protein (PTB) via the chromatin binding protein MRG15, which specifically recognizes trimethylated H3K36. PTB binds to exon-specific silencing elements, resulting in the repression of these exons (Luco et al.,2010). They suggested the existence of several of these "adaptor complexes", which consist of a combination of (1) a modified histone tail, (2) a chromatin binding protein that recognizes this specific histone modification, and (3) a splice regulator that is recruited to the histone mark via the chromatin-binding protein (reviewed by Luco et al., 2011). Recent findings suggest that alternative splicing and trimethylated H3K36 can bi-directionally influence each other. Co-transcriptional spliceosome assembly and splicing enhances the recruitment of SETD2, leading to enrichment of H3K36me3 (De Almeida et al., 2011; Kim et al., 2011). All these findings suggest that SETD2 inactivation results in aberrant splicing, which is common in cancers and plays a role in tumorigenesis and maintenance (Venables et al., 2009).

To assess the effect of *SETD2* inactivation on transcript levels and alternative splice site choice in proximal tubular epithelial cells, we transiently knocked down *SETD2* in these cells, resulting in significantly decreased global H3K36me3 levels.

By RNAseq, we assessed the effect of this knockdown on transcript levels and exon usage genome-wide. As expected from the literature (Edmunds et al., 2008; Yoh et al., 2008), there was a positive correlation between the presence of H3K36me3 and gene expression in the wild-type PTEC sample. A limited set of genes showed a significant change in transcript levels as a result of *SETD2* knockdown. Gene set enrichment analysis of the set of 55 downregulated genes showed a significant enrichment for angiogenesis-related processes, and the ErbB- and P53 signaling pathways. The enrichment for angiogenesis-related genes is in concordance with the crucial role for SETD2 in embryonic vascular remodeling in mice (Hu et al., 2010). SETD2 was previously suggested to play a role in the transcriptional control of a subset of P53 target genes by direct interaction with P53 (Xie et al., 2008). It remains to be explored which of the observed changes in transcript levels are a direct consequence of *SETD2* knockdown, and what mechanisms link SETD2 function and/or H3K36me3 with the regulation of transcript levels.

Considering alternative splicing, our preliminary data analysis indeed revealed an apparent change in alternative splicing as a result of *SETD2* knockdown for a set of genes, which corroborated the suggested role of SETD2 via H3K36me3 in alternative splice site choice for a subset of target genes (Luco et al., 2010). Several of the genes that show a change in alternative exon usage encode proteins that function in processes that are relevant to tumorigenesis, such as WNT signaling (*Axin*), TGF-β signaling (*LTBP3*), cell cycle entry control (*PUM2*), and DNA repair (*RAD1*). Besides genes that show a change in alternative splice site choice, some genes appeared to be transcribed from a different transcriptional start site as a result of *SETD2* knockdown. It was recently shown that KDM5B, a H3K4me3 demethylase, is recruited to H3K36me3 via interaction with the chromodomain protein MRG15. This results in a decrease of intragenic H3K4me3, thereby suppressing cryptic intragenic transcription (Xie et al., 2011). Thus, loss of H3K36me3 could indeed result in alternative transcriptional start site choice. The exact mechanisms and proteins that underlie the observed changes in alternative splicing or transcriptional start site, and the effect of these changes on the function of the encoded proteins, as well as their potential role in ccRCC tumorigenesis, require further functional study.

PBRM1

The observed frequency of inactivating mutations in *SETD2* in ccRCC tumors suggested that there were probably other potential ccRCC tumor suppressor genes located at 3p21 to be discovered. As our GINI analysis did not reveal any other candidates, we decided to make use of the recently developed next-generation sequencing (NGS) technology to look for additional inactivated genes (***Chapter 5***). This resulted in the identification of inactivating mutations in *PBRM1* in seven out of ten ccRCC tumors and inactivating mutations in *BAP1* and *ZNF197* in one tumor. Moreover, we identified inactivating mutations in *PBRM1* in five out of ten ccRCC-derived cell lines. *PBRM1* is located at 3p21, and encodes BAF180, which is the chromatin targeting subunit of the SWI/SNF nucleosome remodeling complex (Thompson et al., 2009). SWI/SNF complexes are large ATP-dependent chromatin-remodeling machines that change

the accessibility of DNA by controlling the dynamics of nucleosome occupancy. The frequent inactivation of *PBRM1* in ccRCC has also been reported by Varela et al. (2011), who identified inactivating mutations in *PBRM1* in 92 out 227 (41%) primary ccRCC tumors. This strongly suggests that *PBRM1* is an important ccRCC tumor suppressor gene that plays a role in tumorigenic development in a large fraction of ccRCC samples. Several other components of the SWI/SNF remodeling complex are involved in the development of different cancer types (Schneppenheim et al., 2010; Rodriguez-Nieto et al., 2011; Jones et al., 2010; reviewed by Reisman et al., 2009). The identification of inactivating mutations in *PBRM1* in four breast cancer samples (Xia et al., 2008), a pancreatic cancer sample (Jones et al., 2008), and several cell lines derived from different tumor types (Varela et al., 2011) provide additional support for *PBRM1* being a true tumor suppressor gene. Furthermore, Varela et al. (2011) showed that siRNA-mediated knockdown of *PBRM1* in ccRCC-derived cell lines with wild-type *PBRM1* results in a significant increase in proliferation, increased colony formation in soft agar, and increased cell migration. However, more *in vivo* evidence is needed to prove the tumor suppressive role of PBRM1 in ccRCC.

The frequent inactivation of *BAP1* in other tumor types (Bott et al., 2011; Harbour et al., 2010), and the role of the encoded protein in chromatin regulation, make *BAP1* an interesting candidate ccRCC cancer gene. The truncating mutation in *ZNF197*, which was identified in a ccRCC sample in which no *VHL* mutation was found, is interesting, given the fact that an alternative isoform of ZNF197 encodes a protein that is possibly involved in VHL-mediated HIF-1-α transactivation (Li et al., 2003). Inactivation of *ZNF197* might therefore be redundant to *VHL* inactivation. Mutational screening of a larger panel of ccRCC tumors is however needed to give a better estimate of the mutational frequency in ccRCC for these genes.

ccRCC genetics

Inactivation of individual ccRCC tumor suppressor genes such as *VHL, PBRM1*, or *SETD2* could confer growth advantage upon the proximal tubular epithelial cell. However, the observed concomitant mutations in *PBRM1*, *VHL*, and *SETD2* in a subset of ccRCC samples suggest that at least three mutations are required for a proximal tubular epithelial cell to develop into a ccRCC tumor. As shown by Young et al. (2008) in a Cre/loxP mouse model, *VHL* inactivation alone is indeed not sufficient to lead to ccRCC tumors in mice. Generation of single-, double-, as well as triple-conditional knockout mice that represent the possible "combinations" of inactivation of *VHL, PBRM1*, and *SETD2* might therefore be necessary to confirm the tumorigenic effect of inactivation of these genes in proximal tubular epithelial cells. These knockout mice could also provide clues about the possible mechanisms through which inactivation of the individual genes contributes to ccRCC development, as well as the probable order in which the mutational events take place during ccRCC development. One possibility is that inactivation of these genes contributes to ccRCC development by affecting different pathways. Alternatively, the functions of the proteins which are encoded by these genes might converge on the same pathways, and inactivation of these genes might therefore mutually strengthen their effects. PBRM1 plays a role in coordinating

senescence by regulating the transcriptional activity of a subset of p53 target genes (Burrows et al., 2010), and has been suggested to be involved in the cellular response to hypoxia (Kenneth et al., 2009), suggesting a functional overlap with VHL function. Indeed, Varela et al. (2011) showed that nearly all primary ccRCC tumors that harbor inactivating mutations in *PBRM1* have a "hypoxia signature", including those in which no *VHL* mutations were identified.

An extensive cross-talk has been shown to exist between different histone modifications and the enzymes that are responsible for these modifications (Kouzarides et al., 2007). Interestingly, the H3K27me3 demethylase UTX travels with the elongating RNAPOLII complex (Smith et al., 2007). It is therefore tempting to speculate that the inactivation of the different histone methylation enzymes observed in ccRCC tumors affects the same set of target genes. Recently, VHL has been shown to regulate gene expression and tumor growth through the regulation of histone demethylases (Beyer et al., 2008; Niu et al., 2011; Yang et al., 2009). We therefore propose that inactivation of *VHL* has a tumorigenic effect in PTECs partially as a result of deregulation of histone demethylases. Inactivating mutations in histone methylation enzymes such as *SETD2* irreversibly alter histone modification patterns, resulting in a loss of cellular identity, and contributing to ccRCC development (*Figure* 1).

In conclusion, our data indicate that the short arm of chromosome 3 harbors several genes that could contribute to ccRCC development when inactivated (*Figure 2).*

The frequently observed 3p loss in ccRCC tumors could be driven by the inactivation of different ccRCC tumor suppressor genes located at 3p, or combinations of 3p tumor suppressors. In general, it is assumed that a subtle mutation in one allele of a tumor suppressor gene provides a "first hit", after which loss of a larger chromosomal segment results in loss of the other allele, and therefore in the functional inactivation of the tumor suppressor gene. Given the apparent presence of multiple candidate ccRCC tumor suppressors on 3p, it is however tempting to speculate that 3p loss alone could already confer growth advantage upon the proximal tubular epithelial cell, and could therefore provide an early step in tumorigenic development.

Several of the mutated genes in ccRCC encode proteins that modify histone tails or remodel chromatin, suggesting that aberrant chromatin regulation is a core driver event in the development of clear cell renal cell carcinoma. Unraveling the mechanisms through which aberrant chromatin regulation contributes to ccRCC development will be an important area of future ccRCC research, and may result in the identification of novel therapeutic targets.

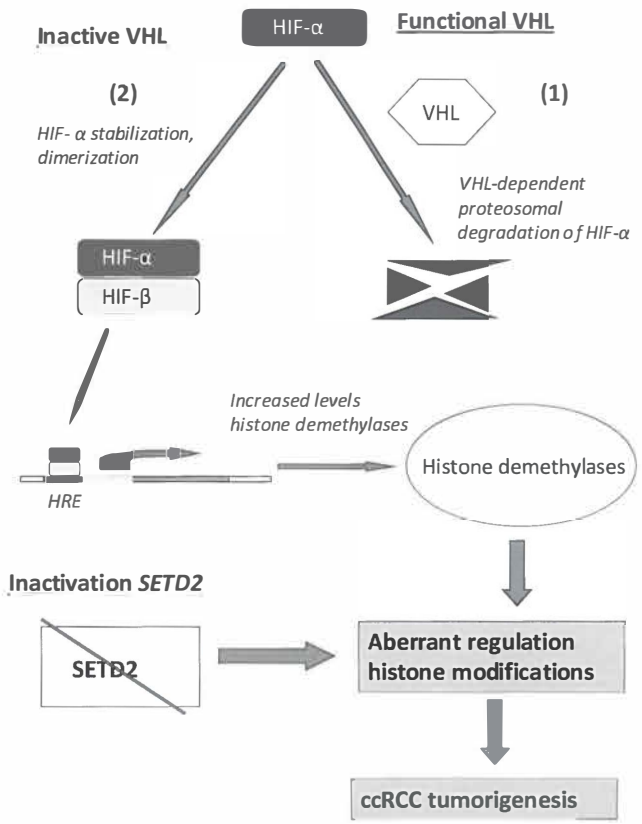


Figure 1. Role of VHL in the regulation of histone modifications. (1) VHL forms part of an E3-ubiquitin ligase complex that targets the transcription factor Hypoxia inducible factor (HIF)-α for degradation. **(2)** Inactivation of *VHL* results in the accumulation of HIF-α, which forms a heterodimer with HIF-β, and translocates to the nucleus, where it binds to hypoxia-responsive elements (HRE) and activates the transcription of target genes. Recently, these target genes have been shown to include several histone demethylases (Beyer et al., 2008; Niu et al., 2008). Inactivation of *VHL* therefore results in consistently high histone demethylase levels, resulting in the aberrant regulation of histone modifications, with possibly tumorigenic consequences. Inactivating mutations in histone methyltransferase *SETD2* result in total loss of H3K36me3, irreversibly altering histone modification patterns. The tumorigenic effects of *VHL* inactivation and inactivation of *SETD2* and other histone modifying enzymes that have been shown to be inactivated in a subset of ccRCCs (Dalgliesh et al., 2010) might therefore converge, at least partially, on the same pathways.

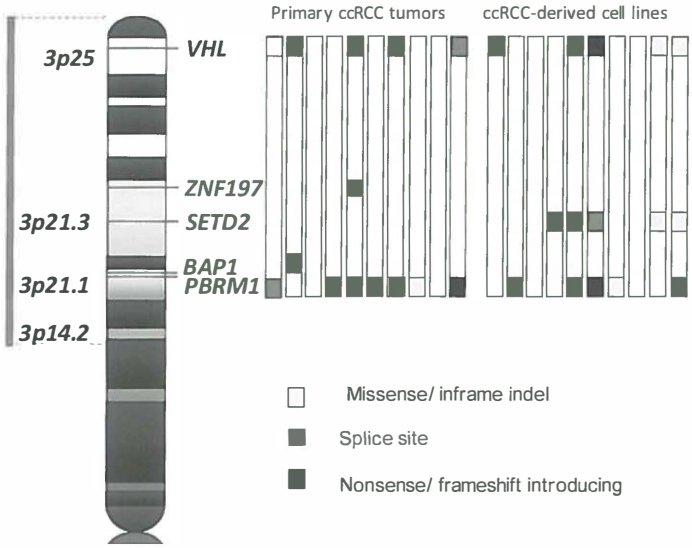


Figure 2. Schematic picture of the short arm of chromosome 3 (3p), showing the locations of "3p genes" in which inactivating mutations were identified during this project. The columns represent the ten ccRCC tumor samples that were screened for mutations in 3p genes by targeted exome-sequencing (Chapter 5) and the ten ccRCC-derived cell lines. These cell lines were screened for *VHL*, *PBRM1*, and *SETD2* mutations, but not for mutations in *BAP1* and *ZNF197*. The vertical line represents the minimally deleted region among this set of ccRCC samples, comprising a large terminal part of 3p. The little squares represent the identified mutations. Targeted 3p exome sequencing of ten primary ccRCC did not reveal any inactivating mutations in *SETD2*. However, screening for *SETD2* mutations in another set of primary ccRCC tumors revealed missense mutations in two out of ten samples (Chapter 2).

REFERENCES

Al Sarakbi W, Sasi W, Jiang WG, et al. The mRNA expression of SETD2 in human breast cancer: correlation with clinico-pathological parameters. BMC Cancer 2009;9:290.

Andersson R, Enroth S, Rada-Iglesias A, et al. Genome Res 2009;19:1732.

Baldewijns MM. Van Vlodrop IJH, Vermeulen PB, et al. VHL and HIF signalling in renal cell carcinogenesis. Journal of Pathology 2010;221:125.

Beyer S, Kristensen MM, Jensen KS, et al. The histone demethylases JMJD1A and JMJD2B are transcriptional targets of hypoxia-inducible factor HIF. J Biol Chem 2008;283:36542.

Bott M, Brevet M, Taylor BS, et al. The nuclear deubuquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. Nature Genetics 2011;43:668.

Braga E., Senchenko V, Bazov I, et al. Critical tumor-suppressor gene regions on chromosome 3P in major human epithelial malignancies: allelotyping and quantitative real-time PCR. Int J Cancer 2002;100:534.

Burrows AE, Smogorzewska A, Elledge SJ. Polybromo- associated BRG-associated factor components BRD7 and BAF180 are critical regulators of p53 required for induction of replicative senescence. PNAS USA 2010;107:14280.

Chen CZ. MicroRNAs as oncogenes and tumor suppressors. N Engl J Med 2005;353:1757.

Chudek J, Wilhelm M, Bugert P, et al. Detailed microsatellite analysis of chromosome 3p region in non-papillary renal cell carcinomas. Int J Cancer 1997;73:225.

Clark PE. The role of VHL in clear-cell renal cell carcinoma and its relation to targeted therapy. Kidney International 2009;76:939.

Dalgliesh GL, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying enzymes. Nature 2010;463:360.

De Almeida SF, Grosso AR, Koch F, et al. Splicing enhances recruitment of methyltransferase HYPB/SETD2 and methylation of histone H3 Lys 36. Nat Struct Mol Biol 2011;18:83.

De la Mata M, Alonso CR, Kadener S, et al. A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell 2003;12:525.

Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/ SETD2 mediates all H3K36 trimethylation. EMBO J 2008;27:406.

Füllgrabe J, Kavanagh E, Joseph B. Histone onco-modifications. Oncogene 2011;30:3391.

Garzon R, Fabbri M, Cimmino A, et al. MicroRNA expression and function in cancer. Trends Mol Med 2006;12:580.

Gibbons DL, Lin W, Creighton CJ, et al. Contextual extracellular cues promote tumor cell EMT and metastasis by regulating miR-200 family expression. Genes & Dev 2009;23:2140.

Gnarra JR, Tory K, Weng Y, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. Nature Genetics 1994:7;85.

Gregory PA, Bert AG, Paterson EL, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nature Cell Biology 2008;10:593.

Harbour JW, Onken MD, Roberson ED, et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. Science 2010;330:1410.

Harten SK, Shukla D, Barod R, et al. Regulation of renal epithelial tight junctions by the von Hippel-Lindau tumor suppressor gene involves occludin and claudin and is independent of E-cadherin. Mol Biol Cell 2009:20;1089.

Hornsby C, Page KM, Tomlinson IP. What can we learn from the population incidence of cancer? Armitage and Doll revisited. Lancet Oncol 2007;8:1030.

Hu M, Sun X-J, Zhang Y-L, et al. Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. Proc Natl Acad Sci U S A 2010;107:2956.

Huff JT, Plocik AM, Guthre C, et al. KR Reciprocal intronic and exonic histone modification regions in humans. Nature Structural & Molecular Biology 2010:17:1495.

Jones S, Wang TL, Shih IeM, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science 2010;330:228.

Jones S, Zhang X, Parsons DW, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 2008;321:1801.

Kaelin Jr. WG. The von Hippel-Lindau tumour suppressor protein: O_2 sensing and cancer. Nature Reviews Cancer 2008:8;865.

Kenneth NS, Mudie S, Van Uden P, et al. SWI/SNF regulates the cellular response to hypoxia. J Biol Chem 2009;284:4123.

Kent OA, Mendell JT. A small piece in the cancer puzzle: microRNAs as tumor suppressors and oncogenes. Oncogene 2006;25:6188.

Kim S, Kim H, Fong N, Erickson B, et al. Pre-mRNA splicing is a determinant of histone H3K36me3 methylation. Proc Natl Acad Sci U S A 2011;108:13563.

Kolasinska-Zwierz P, Down T, Latorre I, et al. Differential chromatin marking of introns and expressed exons by H3K36me3. Nature Genetics 2009;41:376.

Kouzarides T. Chromatin Modifications and their function. Cell 2007;128:693.

Li Z, Wang D, Na X, et al. The VHL protein recruits a novel KRAB-A domain protein to repress HIF-1alpha transcriptional activity. EMBO J 2003;22:1857.

Luco RF, Allo M, Schor IE, et al. Epigenetics in alternative pre-mRNA splicing. Cell 2011; 144:16.

Luco RF, Pan Q, Tominaga K, et al. Regulation of alternative splicing by histone modifications. Science 2010;327:996.

Mandriota SJ, Turner KJ, Davies DR, et al. HIF activation identifies early lesions in VHL kidneys: evidence for site specific tumor suppressor function in the nephron. Cancer Cell 2002;1:459.

Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011;476:298.

Nickerson ML, Jaeger E, Shi Y, et al. Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. Clin Cancer Res 2008;14:4726.

Niu X, Zhang T, Liao L, et al. The von Hippel-Lindau tumor suppressor protein regulates gene expression and tumor growth through histone demethylase JARID1C. Oncogene 2011. [Epub ahead of print].

Reisman D, Glaros S, Thompson EA. The SWI/SNF complex and cancer. Oncogene 1009;28:1653.

Rodriguez-Nieto S, Canada A, Pros E, et al. Massive parallel DNA pyrosequencing analysis

of the tumor suppressor BRG1/SMARCA4 in lung primary tumors. Hum Mutat 2011;32: 1999.

Schneppenheim R, Frühwald MC, Gesk S, et al. Germline mutation and somatic inactivation of *SMARCA4/BRG1* in a family with rhabdoid tumor predisposition syndrome. Am J Hum Genet 2010;86:279.

Schor IE, Rascovan N, Pelisch F, et al. Neuronal cell depolarisation induces intragenic chromatin modifications affecting NCAM alternative splicing. Proc Natl Acad Sci U S A 2009;106:4325.

Smith ER, Lee MG, Winter B, et al. *Drosophila* UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. Mol Cell Biol 2008;28:1041.

Spies N, Nielsen CB, Padgett RA, et al. Biased chromatin signatures around polyadenylation sites and exons. Mol Cell 2009;36:245.

Starr TK, Allaei R, Silverstein KA,et al. A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. Science 2009;323:1747.

Sükösd F, Kuroda N, Beothe T, et al. Deletion of chromosome 3p14.2-p25 involving the VHL and FHIt genes in conventional renal cell carcinoma. Cancer Res 2003;63:455.

Thoenes W, Störkel St, Rumpelt HJ. Histopathology and classification of renal cell tumors (adenomas, oncocytomas, and carcinomas): the basic cytological and histopathological elements and their use for diagnostics. Path Res Pract 1985;181:125.

Thompson M. Polybromo-1: the chromatin targeting subunit of the PBAF complex. Biochimie 2009;91:309.

Toma MI, Grosser M, Herr A, et al. Loss of heterozygosity and copy number abnormality in clear cell renal cell carcinoma discovered by high-density affymetrix 10K single nucleotide polymorphism mapping array. Neoplasia 2008;10:634.

Van den Berg A, Dijkhuizen T, Draaijers TG, et al. Analysis of multiple renal cell adenomas and carcinomas suggests allelic loss at 3p21 to be a prerequisite for malignant development. Genes Chromosomes Cancer 1997;19:228.

Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 2011;469:539.

Venables JP. Aberrant and alternative splicing in cancer. Cancer Res 2004;64:7647.

Venables JP, Klinck R, Koh C, et al. Cancer-associated regulation of alternative splicing. Nat Struct Mol Biol 2009;16:670.

Xia X, Lemieux ME, Li W, et al. Integrative analysis of HIF binding and transactivation reveals its role in maintaining histone methylation homeostasis. Proc Natl Acad Sci U S A 2009;106:4260.

Xia W, Nagase S, Montia AG, et al. BAF180 is a critical regulator of p21 induction and a tumor suppressor mutated in breast cancer. Cancer Res 2008;68:1667.

Xie L, Pelz C, Wang W, Bashar A, et al. KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. EMBO J 2011;30:1473.

Xie P, Tian C, An L, et al. Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. Cell Signal 2008;20:1671.

Yang J, Ledaki I, Turley H, et al. Role of hypoxia-inducible factors in epigenetic regulation via histone demethylases. Ann NY Acad Sci 2009;1177:185.

Yoh SM, Lucas JS, Jones KA. The Iws:Spt6: CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev 2008;22:3422.

Young AP, Schlisio S, Minamishima YA, et al. VHL loss actuates a HIF-independent senescence programme mediated by Rb and p400. Nature Cell Biology 2008;10:361.

SUMMARY

Renal cell carcinoma (RCC) is the most common type of kidney cancer in adults, and accounts for 3% of all new cancers diagnosed in the Western world. Clear cell renal cell carcinoma (ccRCC) is by far the most prevalent subtype, comprising more than 70% of adult malignant kidney tumors. ccRCC tumor are thought to be derived from the epithelial cells of the proximal tubuli (PTECs).

In most cases, ccRCC tumors are still localized at the moment they are found. Surgical removal is the standard of care for these tumors, and offers long-term recurrence-free survival, although for patients who suffer from metastatic disease or develop distant relapse, the scenario is worse. Renal cell carcinomas are resistant to traditional cytotoxic chemotherapy. Immunotherapy, which has been the most frequently used therapy until recently, has not been satisfactory in most cases either.

Advances in the understanding of ccRCC biology have led to the development of approaches that target pathways that are characteristic and essential for ccRCC tumors, and show superior response rates compared to immunotherapy.

At the genomic level, ccRCC is characterized by a range of copy number alterations, of which loss of the short arm of chromosome 3 is by far the most frequent one. Genomic regions with a high incidence of allelic loss are thought to harbor so-called tumor suppressor genes (TSGs). The *Von Hippel Lindau* (*VHL*) gene, located at 3p25, is a known TSG and is bi-allelicely inactivated in a large fraction of ccRCC tumors. However, *VHL* is not inactivated in all ccRCC tumors, and inactivation of *VHL* alone does not result in ccRCC development, suggesting that additional and/or alternative genes play a role in ccRCC development.

The identification of novel TSGs was hampered by the omission of methods to screen large numbers of genes in parallel for small (point) mutations. This situation has changed in the last decade. New methods like Gene Identification by nonsense-mediated mRNA decay Inhibition (GINI), and, more recently, next-generation sequencing-based approaches, have given us new and powerful tools to identify tumor suppressor genes.

This thesis presents the search for novel ccRCC tumor suppressor genes involved in the development of ccRCC.

In Chapter 1, an overview of the current knowledge regarding the epidemiology, etiology, and genetics of ccRCC is presented.

In Chapter 2, we describe the use of GINI in the search for novel ccRCC tumor suppressor genes. In this method, the NMD pathway, a pathway that degrades mRNA molecules that harbor a premature termination codon (PTC), is pharmacologically inhibited in cell lines. This results in the accumulation of transcripts harboring a PTC. The increased transcript levels of these mRNAs can be detected using gene expression microarrays. The presence of the mutation in this transcript causing the PTC still needs to be identified by Sanger sequencing. We applied a GINI strategy on a panel of ten ccRCC-derived cell lines and combined several methods and criteria to maximize the chances of finding ccRCC tumor suppressor genes and, at the same time, minimize the number of false-positive candidates that hampered the analysis of GINI strategies described earlier. This led to the identification of mutations introducing a premature termination codon in histone methyltransferase gene *SETD2*,

located at 3p21, in two cell lines. Sequencing of *SETD2* in the other cell lines revealed inactivating mutations in three of them. SETD2 is responsible for trimethylation of lysine 36 of histone H3 (H3K36me3). Consistent with this function, we observed a loss or a decrease of H3K36me3 in seven out of the ten ccRCC cell lines. Identification of missense mutations in 2 out of 10 primary ccRCC tumor samples added support to the involvement of loss of SETD2 function in the development of ccRCC tumors.

In Chapter 3, we evaluate the use of GINI in the search for TSGs by using the current knowledge on ccRCC genetics. We show that GINI has some intrinsic limitations. TSGs that code for relatively small proteins and/or have only a limited number of exons appear to have a limited chance on being identified by GINI. Moreover, the succes of the method is dependent on the quality of the used gene expression platform, and the availability of tumor-derived cell lines.

In Chapter 4, we describe the initial steps that should result in the elucidation of the mechanisms through which *SETD2* inactivation contributes to ccRCC development. H3K36me3 has been suggested to have a function in alternative splice site choice. Misregulation of splicing is common in cancer and plays a role in tumor development. We knocked down *SETD2* in proximal tubular epithelial cells and indeed observed an apparent change in alternative splicing for a set of genes, several of which are interesting with respect to tumorigenic development. Moreover, a set of genes showed a significant change in transcript level as a result of *SETD2* knockdown. The set of downregulated genes showed a significant enrichment for genes that function in angiogenesis-related processes and genes that are involved in the p53-pathway. This is in concordance with the earlier described role of SETD2 in embryonic vascular remodeling and its function in the regulation of a subset of P53 target genes.

The recently developed next-generation sequencing (NGS) technologies make it possible to screen large numbers of genes in parallel for mutations. We continued our search for ccRCC tumor suppressor genes by sequencing all coding regions and exon-intron boundaries from the short arm of chromosome 3 in ten primary ccRCC tumors, using targeted exome capturing followed by NGS, as described in chapter 5. This led to the identification of inactivating mutations in *PBRM1* in seven out of ten ccRCC tumors and inactivating mutations in *BAP1* and *ZNF197* in one tumor each. Moreover, we identified inactivating mutations in *PBRM1* in five out of ten ccRCC-derived cell lines. *PBRM1* is located at 3p21, and encodes BAF180, which is the chromatin targeting subunit of the SWI/SNF nucleosome remodeling complex. SWI/SNF complexes are large ATP-dependent chromatin-remodeling machines that change the accessibility of DNA by controlling the dynamics of nucleosome occupancy. The observed mutational frequency of *PBRM1* suggests that this is a major ccRCC tumor suppressor gene. Several ccRCC samples show concomitant mutations in *VHL* and *PBRM1*, suggesting that inactivation of these genes is functionally nonredundant.

Besides protein coding genes, non-coding transcripts, including miRNAs, have a function in the development of several tumor types. In Chapter 6, we compared miRNA transcript levels in ccRCC-derived cell lines with the miRNA levels in normal proximal tubular epithelium, and correlated these transcript levels with copy numbers and transcript levels of predicted targets. Among the significantly downregulated

miRNAs in the ccRCC cell lines were all miRNA-200 family members and miR-205. As a result of loss of the miR-200 family and miR-205, epithelial cells undergo an epithelial-to-mesenchymal transition (EMT), which, in a tumorigenic context, confers a metastatic potential to the cells. Thus, our data suggest a role for these miRNAs in ccRCC progression.

In Chapter 7, the results described in this thesis are discussed. The short arm of chromosome 3 appears to harbor several genes that could contribute to ccRCC development when inactivated. Our data give new insights into the mechanisms that are involved in the development of ccRCC. Several of the mutated genes in ccRCC encode proteins that modify histone tails or remodel chromatin, suggesting that aberrant chromatin regulation is a core driver event in the development of clear cell renal cell carcinoma.

NEDERLANDSE SAMENVATTING

Niercelcarcinoom is het meest voorkomende type nierkanker en neemt zo'n drie procent van alle gevallen van kanker bij volwassenen voor zijn rekening. Verreweg de meeste niercelcarcinomen (meer dan 70 procent) zijn van het "clear cell" subtype.

Clear cell niercelcarcinomen (ccRCCs) ontstaan uit de epitheelcellen van de proximale tubuli en zijn in de meeste gevallen nog niet uitgezaaid op het moment dat ze worden gevonden. De standaardbehandeling voor een patiënt met een ccRCC is de chirurgische verwijdering van de tumor. Voor patiënten met een gelokaliseerde tumor biedt deze behandeling langdurige overlevingskansen. Voor patiënten die te maken hebben met uitzaaiingen is het scenario echter veel minder gunstig. ccRCCs zijn resistent tegen traditionele chemotherapie. Ook immunotherapie, een behandeling die tot voor kort veel werd toegepast, slaat in de meeste gevallen niet aan. De toegenomen kennis en inzicht in de biologie van ccRCC hebben recentelijk geleid tot de ontwikkeling van behandelingen die aangrijpen op 'pathways' die karakteristiek en essentieel zijn voor ccRCC. Vergeleken met immunotherapie hebben deze gerichte therapieën een veel hogere 'response rate'.

ccRCCs worden genetisch gezien gekenmerkt door een aantal chromosomale veranderingen, waarbij verlies van de korte arm van chromosoom 3 het meest voorkomt. Het veelvuldige verlies van 3p in ccRCCs is een aanwijzing voor de aanwezigheid van tumorsuppressor- genen in dit genomische segment. Het *Von Hippel Lindau* (*VHL*) gen, gelegen op 3p25, is een bekend ccRCC tumorsuppressor-gen en is bi- allelisch geïnactiveerd in de meerderheid van ccRCC tumoren. *VHL* is echter niet in alle ccRCC tumoren geïnactiveerd en inactivatie van *VHL* alleen leidt niet tot de ontwikkeling van ccRCC tumoren. Vermoedelijk spelen additionele en/ of alternatieve genen dan ook een rol in de ontwikkeling van ccRCC.

De zoektocht naar nieuwe tumorsuppressor- genen werd lang bemoeilijkt door de afwezigheid van methodes waarmee grote aantallen genen tegelijk kunnen worden gescreend voor kleine (punt)mutaties. Deze situatie is het in het laatste decenium veranderd. Methodes zoals gene identification by nonsense-mediated mRNA decay inhibition (GINI) en "next- generation sequencing" technieken hebben ons nieuwe krachtige instrumenten gegeven om tumorsuppressor- genen te identificeren.

In dit proefschrift beschrijven we de zoektocht naar nieuwe tumorsuppressor-genen die betrokken zijn bij de ontwikkeling van ccRCC.

In *Hoofdstuk 1* wordt een overzicht gegeven van de huidige kennis betreffende de epidemiologie, etiologie en genetica van ccRCC.

In *Hoofdstuk* 2 beschrijven we het gebruik van GINI in de zoektocht naar nieuwe ccRCC tumorsuppressor- genen. In deze methode wordt de nonsense- mediated mRNA decay pathway geblokkeerd in cellijnen. Deze pathway ruimt mRNA moleculen die ten gevolge van mutaties een prematuur stopcodon (PTC) hebben. Uitschakeling van deze pathway resulteert in de ophoping van transcripten met een PTC. De ophoping van deze transcripten kan worden gedetecteerd met behulp van gen-expressie microarrays. Sanger sequencing wordt vervolgens gebruikt om de aanwezigheid van een PTC in de betreffende transcripten aan te tonen.

We hebben een GINI- strategie toegepast op een panel van tien ccRCC- cellijnen. Verschillende methodes en criteria werden daarbij toegepast om de kans op het

vinden van nieuwe tumorsuppressor-genen te vergroten en tegelijkertijd het aantal vals- positieve kandidaten te beperken. Dit resulteerde in de identificatie van PTC-introducerende mutaties in het histon- methyltransferase- gen *SETD2*, gelegen op 3p21, in twee cellijnen. Sequencing van *SETD2* in de overige cellijnen resulteerde in de identificatie van inactiverende mutaties in nog drie cellijnen. SETD2 is verantwoordelijk voor trimethylering van lysine 36 van histon 3 (H3K36me3). De bevinding dat zeven van de tien ccRCC cellijnen een verlies of in ieder geval een significante afname van H3K36me3 laten zien is in lijn met de betrokkenheid van *SETD2* functie in de ontwikkeling van ccRCC. Extra bewijs voor de betrokkenheid van verlies van SETD2-functie in de ontwikkeling van ccRCC- tumoren werd geleverd door de identificatie van missense mutaties in *SETD2* in twee van de tien gescreende primaire ccRCC- tumoren.

In *Hoofdstuk* 3 evalueren we het gebruik van GINI in de zoektocht naar tumorsuppressor- genen. Hierbij hebben we gebruik gemaakt van de huidige kennis over de genetica van ccRCC. We laten zien dat GINI enkele intrinsieke beperkingen heeft. Tumorsuppressor- genen die coderen voor relatief kleine eiwitten en/ of een beperkt aantal exonen hebben blijken slechts een kleine kans te hebben om geïdentificeerd te worden met behulp van GINI. Daarnaast is de kans op succes afhankelijk van de kwaliteit van het gebruikte expressie- platform.

In *Hoofdstuk 4* beschrijven we de initiële stappen die moeten leiden tot de ontrafeling van de mechanismes waardoor inactivatie van *SETD2* bijdraagt aan de ontwikkeling van ccRCC. H3K36me3 is betrokken bij de keuze van alternatieve splice sites. Misregulatie van splicing komt vaak voor in kanker en speelt een rol bij tumor-ontwikkeling. Inactivitatie van *SETD2* draagt daarom mogelijk bij aan de ontwikkeling van ccRCC door een verandering in splicing van targetgenen. Met behulp van siRNA zorgden wij voor een afname van de hoeveelheid SETD2 in proximale tubulaire epitheelcellen en observeerden inderdaad een verandering in alternatieve splicing voor een aantal genen. Verschillende van deze genen zijn goed te koppelen aan tumorontwikkeling. Daarnaast laat een aantal genen als gevolg van de afname van SETD2 een verandering in transcript- niveau zien. De groep downgereguleerde genen is significant verrijkt voor genen die functioneren in angiogenese- gerelateerde processen en voor genen die een rol spelen in de p53 pathway. Dit is in overeenstemming met de eerder beschreven rol van SETD2 in embryonale vasculaire modellering, en de rol van SETD2 in de regulatie van een subset van P53 targetgenen.

De recent ontwikkelde "next generation" sequencing (NGS) technieken maken het mogelijk een groot aantal genen in parallel te screenen voor mutaties. In *Hoofdstuk 5* zetten wij onze zoektocht naar nieuwe ccRCC tumorsuppressor- genen voort door alle coderende regio's en exon-intron grenzen van de korte arm van chromosoom 3 in tien primaire ccRCC tumoren te sequencen met behulp van 'targeted exome capturing' gevolgd door NGS. Dit resulteerde in de identificatie van inactiverende mutaties in *PBRM1* in zeven van de tien ccRCC tumoren, en inactiverende mutaties in *BAP1* en *ZNF197* ieder in één tumor. Daarnaast identificeerden we inactiverende mutaties in *PBRM1* in vijf van de tien ccRCC cellijnen. *PBRM1* ligt op 3p21, en codeert voor BAF180, een onderdeel van het SWI/SNF complex. SWI/SNF- complexen zijn grote ATP- afhankelijke chromatine- modellerende machines die de toegankelijkheid

van DNA kunnen beïnvloeden door de positie van nucleosomen langs het DNA te veranderen. De hoge frequentie van mutaties in *PBRM1* in ccRCC tumoren doet vermoeden dat dit een belangrijk ccRCC tumorsuppressor- gen is.

Behalve genen die coderen voor eiwitten spelen ook niet- coderende transcripten zoals microRNAs een rol in de ontwikkeling van verschillende tumortypes. In *Hoofdstuk 6* vergelijken we de miRNA niveaus in ccRCC- cellijnen met de miRNA niveaus in normaal proximaal tubulair epitheel, en correleren de hoeveelheid transcripten met het aantal genomische kopieën en de transcript- niveaus van voorspelde mRNA targets. Onder de significant downgereguleerde microRNAs in de ccRCC-cellijnen bevonden zich alle leden van de miR-200 familie en miR-205. Als gevolg van verlies van functie van deze microRNAs ondergaan cellen een zogenaamde epitheel- mesenchym transitie (EMT), wat in een tumorigene context bijdraagt aan het metastatisch potentieel van de cel. De miR-200 familie en miR-205 lijken dus een rol te spelen in de progressie van ccRCC.

In *Hoofdstuk 7* worden de resultaten die zijn beschreven in dit proefschrift bediscussieerd. De korte arm van chromosoom 3 blijkt verschillende genen te bevatten die mogelijk een rol spelen bij aan de ontwikkeling van ccRCC. Onze data geven nieuwe inzichten in de mechanismes die betrokken zijn bij de ontwikkeling van ccRCC. Verschillende van de gemuteerde genen in ccRCC coderen voor eiwitten die een rol spelen in het modificeren van histonen en het modelleren van chromatin. Een verstoring van de regulatie van chromatine lijkt dan ook een essentiële gebeurtenis in de ontwikkeling van clear cell renal cell carcinoma.

ACKNOWLEDGEMENTS / DANKWOORD

Dit proefschrift is het resultaat van bijna vier-en-een-half jaar onderzoek aan de afdeling Genetica. Vier-en-een-half jaar lijkt lang, maar, cliché en dus waar: de tijd vliegt en voor je het weet ben je bezig met het schrijven van een dankwoord. Terugkijkend kan ik zeggen dat het een mooie en leerzame periode is geweest. Graag wil ik dan ook de collega's bedanken met wie ik tijdens deze periode te maken heb gehad en die- direct of indirect- een rol hebben gespeeld bij het tot stand komen van dit proefschrift.

Allereerst gaat mijn dank uit naar mijn begeleiders, mijn co-promotores Dr. Klaas Kok en Dr. Eva van den Berg - de Ruiter, en mijn promotor Prof. Dr. Robert M. W. Hofstra.

Klaas, bedankt voor de mogelijkheid om mijn promotie-onderzoek onder jouw begeleiding te kunnen doen. Ik heb je manier van begeleiden, met je nuchtere kijk en droge humor, als zeer prettig ervaren. Je hebt me geleerd om zelfstandig onderzoek te doen en je was altijd beschikbaar voor deskundig advies. Het heeft even geduurd voordat we ons eerste echte resultaat hadden, maar ik heb van je geleerd dat geduld en vertrouwen in je eigen data belangrijke eigenschappen zijn om een onderzoek tot een goed einde te kunnen brengen. Uiteraard was het ook een voorrecht om geregeld te kunnen genieten van de gastvrijheid en de kookkunsten van jou en Anke.

Robert, bedankt voor je input in het nierkankerproject. Je duidelijke visie en doortastendheid hebben een belangrijke rol gespeeld bij het tot stand komen van dit proefschrift. Daarbij zijn er volgens mij weinig professoren die op zo'n relaxte manier met hun PhD- studenten omgaan.

Eva, bedankt voor het delen van je kennis, met name op het gebied van de cytogenetica, tijdens onze werkbesprekingen.

I would like to thank the reading committee, formed by Prof. J. M. Nijman, Prof. A. H. M. Geurts van Kessel, and Prof. S. J. M. Jones, for their willingness to read my thesis and for giving their approval for the defense.

Inge, na een half jaar kwam jij als analist onze (zeer) kleine "nierkankergroep" versterken. Ik heb vanaf het begin met erg veel plezier met je samengewerkt. Je inzet, betrokkenheid en goede praktische werk waren cruciaal voor het tot stand komen van het SETD2 paper, en daarmee het hele proefschrift. Een bedankje is hier dus zeer op zijn plaats. Dat je de afdeling "voortijdig" verliet om met aardappels aan de slag te gaan spijtte zowel ons als jouzelf: een goed teken lijkt mij.

Jantine, jij ging daarna als analist aan de slag op het ChIPseq-project. Na een paar jaar Italië weer terug in het lab en in Stad. Het moet even wennen zijn geweest, maar daar heb ik vrij weinig van gemerkt. Vanaf het begin leek het je vanzelf te gaan en kwamen er mooie resultaten. Dat je daarnaast ook zeer plezierig in de omgang bent maakt je een erg fijne collega om mee samen te werken.

De studenten die stage hebben gelopen op het project: bedankt voor al jullie werk en inzet. Matthijs voor het toepassen van GINI op longtumor-cellijnen, Harrie voor het onderzoeken van verschillende histon- modificaties in niertumoren, Cor en Angela voor het sequencen van tumorsamples.

De wekelijkse oncogenetica-werkbesprekingen op woensdagochtend waren, met naast onze "nierkankergroep" onder andere Helga, Renée, Rolf, Jan O en Robert, erg nuttig en bovendien gezellig.

Gerard te Meerman, bedankt voor de nuttige en bij vlagen onnavolgbare uiteenzettingen over statistiek en next- generation sequencing.

Anke van den Berg, bedankt voor onder andere het bediscussiëren van onze miRNA- data en je bijdrage aan de miRNA- paper. Andere mensen van de afdeling Pathologie & Medische Biologie die ik wil bedanken voor hun bijdrage aan verschillende hoofdstukken van dit proefschrift: Harry Hollema, Joost Kluiver, Harry van Goor, Lydia Visser, Nancy Halsema en Geert Harms.

Professor Steven M Jones, thank you for the opportunity to visit the Genome Sciences Centre in Vancouver. Ann, Nina and Misha, thanks for your great help with analyzing our ChIP-seq and RNA-seq data. Misha, it was great to visit Vancouver Island with you and Elize.

De heren van Bioké/ Cell Signaling Technology: Koos Kranenborg, Bryan Ricketts, Robert Akkers en Fons Elstrodt, bedankt voor jullie enthousiaste hulp bij het opzetten van de ChIPseq-experimenten en het beschikbaar stellen van materiaal.

Jackie, bedankt voor het snelle en zorgvuldige "editen" van dit proefschrift en verschillende andere manuscripten.

Het secretariaat (Mentje, Ria, Joke, Hélène en Bote), Marina en Edwin: bedankt voor jullie hulp bij allerhande adminstratieve, financiële en computertechnische zaken.

My roommates: Anna, Ana, Peter, Annemieke, Jorieke, Justyna, Helga, Karin, Tjakko, Jihane, and Jinghua, thank you for the great atmosphere at our room!

Helga, als enige "postdoc" op onze kamer was je, wellicht tot vervelens toe, een dankbare vraagbaak. Je droge sarcastische opmerkingen waren zeer vermakelijk.

Anna, het was leuk om te zien hoe snel en op welke manier je je het Nederlands eigen maakte, dat werd gewaardeerd!

Peter, jij was als enige andere man op de kamer en sportkenner bij uitstek (wat ook van pas kwam bij het pubquizzen) een ideale gesprekspartner.

Karin, ik heb veel met je gelachen, mede dankzij jouw directheid en je bijnaam schiep natuurlijk meteen een band. Ondanks dat ik je nooit heb zien voetballen weet ik dat je een uitstekende tackle in de benen hebt.

Annemieke, ik heb genoten van je vrolijke gezelschap, je betrokkenheid en je humor. Je bevlogenheid in je werk was mooi om te zien. Fijn dat je mijn paranimf wilt zijn!

Verder zorgden de verschillende "vaste bewoners" van het lab ervoor dat het ook daar goed vertoeven was: Paul J, Ron, Krista, Mathieu, Ludolf, Monique, Karen, Soesma, Astrid, DJ Miguel en Rutger, bedankt voor de gezelligheid op het lab, jullie hulp, en verschillende hilarische momenten. Jan, bedankt voor je hulp bij onder andere het Sanger- sequencen. De mensen van de "genome analysis facility": Mariska, Jelkje, Bahram en Pieter, die ook passen in bovengenoemd rijtje, wil ik graag bedanken voor de prettige samenwerking en hun praktische bijdrage aan de verschillende "next-generation-sequencing" projecten beschreven in dit proefschrift.

Rest mij nog een aantal collega's van de afdeling te bedanken die nog niet genoemd zijn, maar die hebben bijgedragen aan een mooie tijd, zowel op als naast

het werk: Christine, Yunia, Eva, Renée, Paul vd Z, Maria, Olga, Duco, Barbara, Suzanne, Agata, Gosia, Mats, Rodrigo, Asia, Dineke, Céline, Cleo, Ellen, Noortje, Mitja. Thank you for a pleasant time!

Ivo, bedankt voor je bereidheid om- last minute- een mooie omslag voor dit proefschrift te ontwerpen.

Willem, we kennen elkaar van jongsaf aan, en vanaf het moment dat we elf jaar geleden beiden in Groningen gingen studeren zijn we beste vrienden. Wat mij betreft dan ook logisch dat jij mijn paranimf bent.

Verder wil ik zoals het betaamt in een dergelijk dankwoord als laatste mijn vrienden en familie bedanken, in het bijzonder mijn lieve zus, zwager en zusje: Hilde en Michel (vanuit Ethiopië) en Renske (vanuit waar dan ook ter wereld, maar recentelijk zowaar NL) en mijn lieve ouders: bedankt voor jullie altijd aanwezige steun en belangstelling!

LIST OF ABBREVIATIONS

A:	adenine
ATP:	adenosine triphosphate
AWS:	associated with SET
BAF180:	BRG1-associated factor 180
BAH:	bromo-adjacent homology domain
BAP1:	BRCA1 associated protein-1
BD:	bromo domain
bp:	basepair(s)
BWA:	Burrows-Wheeler aligner
C:	cytosine
ccRCC:	clear cell renal cell carcinoma
cDNA:	complementary DNA
CDF:	cumulative distributive function
CGH:	comparative genomic hybridization
ChIP:	chromatin immunoprecipitation
ChIPseq:	chromatin immunoprecipitation followed by next-generation sequencing
DAVID:	Database for Annotation, Visualization and Integrated Discovery
dbSNP:	the Single Nucleotide Polymorphism database
del:	deletion
DNA:	deoxyribonucleic acid
ECR:	exon-like sequence composition regions
EMT:	epithelial-to mesenchymal transition
EPHB2:	ephrin type-B receptor 2
FC:	fold change
FDR:	false discovery rate
FHIT:	fragile histidine triad
Fs:	frameshift
G:	guanine
GEO:	Gene Expression Omnibus
GINI:	gene identification by NMD inhibition
GO:	Gene Ontology
H3:	Histone 3
H3K27me3:	Histone 3, trimethylated at lysine residue 27
H3K36me2:	Histone 3, dimethylated at lysine residue 36
H3K36me3:	Histone 3, trimethylated at lysine residue 36
H3K4me3:	Histone 3, trimethylated at lysine residue 4
Hg:	human genome
HIFα:	Hypoxia inducible factor-α
HMG:	high mobility group

IGV:	Integrative Genomics Viewer
Ins:	insertion
IP:	interaction with RNAPOLIIA
Kb:	kilobases
KD:	knock down
KEGG:	Kyoto Encyclopedia of Genes and Genomes
LOH:	loss of heterozygosity
Mb:	megabases
MLH1:	mutL homolog 1
mRNA:	messenger RNA
miRNA:	microRNA
MRG15:	MORF-related gene 15
mTOR:	mammalian target of rapamycin
NCBI:	National Center for Biotechnology Information
NGS:	next-generation sequencing
NMD:	nonsense-mediated mRNA decay
PBRM1:	Polybromo 1
PCR:	polymerase chain reaction
PDGF:	platelet derived growth factor
PI3K:	phosphatidylinositol 3-kinase
PTB:	polypyrimidine tract binding protein
PTC:	premature termination codon
PTEC:	proximal tubular epithelial cell
Pter:	terminal end of the short arm of the chromosome
RASSF1:	ras association (RalGDS/AF-6) domain family member 1
RCC:	renal cell carcinoma
RefSeq:	Reference Sequence
RNAPOLIIA:	RNA polymerase 2A
RNA:	ribonucleic acid
RNAseq:	transcriptome profiling using next-generation sequencing
RPKM:	reads per kilobase per million mapped reads
RT:	room temperature
RT-PCR:	real-time polymerase chain reaction
SET:	suppressor of variegation, enhancer of zeste and thrithorax
SETD2:	SET domain containing 2
siRNA:	small inhibiting RNA
SWI/SNF:	switch/ sucrose nonfermentable
T:	thymine
TSG:	tumor suppressor gene
TSS:	transcription start site
UCSC:	University of California, Santa Cruz
UTR:	untranslated region
VEGF:	vascular endothelial growth factor

VHL:	Von Hippel Lindau
WHO:	World Health Organization
WT:	wild-type
ZNF197:	zinc finger protein 197